

# **RECONHECIMENTO DE FACE INVARIANTE A ILUMINAÇÃO BASEADO EM UMA ABORDAGEM SUPERVISIONADA**

Larissa Natália das Virgens Carneiro  
Universidade Federal de Ouro Preto

Dissertação submetida ao  
Instituto de Ciências Exatas e Biológicas  
Universidade Federal de Ouro Preto  
para obtenção do título de Mestre em Ciência da Computação



*Dedico este trabalho a meus pais, Hélio e Rosinete, ambos exemplos de coragem,  
determinação e bondade para minha vida.*



# RECONHECIMENTO DE FACE INVARIANTE A ILUMINAÇÃO BASEADO EM UMA ABORDAGEM SUPERVISIONADA

## Resumo

A crescente relevância dada aos estudos e pesquisas de sistema automáticos de reconhecimento/identificação de faces capazes de identificar indivíduos nas mais diversas situações é devido às várias possibilidades de aplicações tais como sistemas de segurança, bancários, eleitorais e busca por pessoas desaparecidas. Outro fator é a questão da tarefa de reconhecimento não ser trivial devido aos componentes variantes como envelhecimento, uso de óculos, chapéu, maquiagem, variação de aparência e a variação de iluminação. Esta última é um dos maiores desafios dos sistemas de reconhecimento, pois pode ocultar quase todas as características da face. Assim, o presente trabalho propõe um sistema de reconhecimento de faces invariante à iluminação. O mesmo utiliza como pré-processamento das imagens as técnicas *Local Contrast Enhancement* (LCE) ou normalização da iluminação no domínio *Discrete Cosine Transform* (DCT), na segunda fase é utilizado o DCT para extração de características e na terceira o *Discrimination Power Analysis* (DPA) é usado para redução de dimensionalidade. O reconhecimento é feito com o *Support Vector Machine* (SVM) e os experimentos são realizados em duas etapas. Na primeira são utilizadas as bases de dados Pie e Yale B e o modelo proposto é avaliado quanto ao quesito de variação de iluminação. Na segunda fase são utilizadas as bases JAFFE, AT&T, UMIST e Georgia e o modelo é avaliado quanto à robustez em relação a variação de expressão, rotação e fundo. O método proposto apresenta melhor desempenho e melhores resultados para as variações existentes nas bases testadas.



# RECONHECIMENTO DE FACE INVARIANTE A ILUMINAÇÃO BASEADO EM UMA ABORDAGEM SUPERVISIONADA

## Abstract

The increasing relevance attributed to the field of automatic faces recognition/identification, that can identify people in several situations, is due to several potential applications such as security systems, banking, electoral, and search for missing people. Another factor of interest in this field is that the recognition task is nontrivial. There are several components that can influence the task as aging, the use of glasses, hats, make up and the changes in lighting. The lighting changes are one of the major challenges for the recognition systems, since it can hide almost all the features of the face. This work presents a face recognition method invariant to illumination. To the preprocessing of the images the method uses the technique Local Contrast Enhancement (LCE) or the normalization of the lighting in the Discrete Cosine Transform (DCT) domain. The DCT was also used to feature extraction and the DPA was used to dimensionality reduction. The recognition task is handled with the Support Vector Machine (SVM) and the experiments are carried out in two steps. In the first one two databases Pie and Yale B are used and then the proposed model is evaluated with respect to variation of illumination. In the second step are used the databases JAFFE, ATT, UMIST and Georgia and the model is evaluated with respect to the robustness of changes of facial expressions, rotation and background. The proposed method achieved better performance and results in the tested databases.





## Declaração

Esta dissertação é resultado de meu próprio trabalho, exceto onde referência explícita é feita ao trabalho de outros, e não foi submetida para outra qualificação nesta nem em outra universidade.

Larissa Natália das Virgens Carneiro



## Agradecimentos

Primeiramente agradeço a Deus por me proporcionar a oportunidade de seguir com os meus estudos. Em segundo lugar, aos meus pais pelo carinho, amor, dedicação e incentivo assim como aos meus irmãos, em especial a minha irmã que me ensinou neste tempo o verdadeiro significado desta palavra. A minha sobrinha pelo carinho e por participar de muitos momentos apenas me alegrando.

Ao meu prezado orientador Guillermo Cámara Chávez, por todo conhecimento compartilhado, pela amizade e principalmente, pela compreensão nos momentos que mais precisei.

Agradeço a todos velhos e novos amigos por todo carinho, apoio e as noites intermináveis de estudos e em especial a República Virakopos por me oferecer um lar fora de casa e por muitas vezes se fazer de minha família. Agradeço a Mathias Brito, pelo imenso incentivo e apoio para que eu buscasse e persistisse nessa conquista.

Por fim, agradeço a todos os funcionários e professores do DECOM/UFOP pela atenção, dedicação e por me oferecer uma formação técnica e humana de qualidade da qual tenho muito orgulho.



# Sumário

<b>Lista de Figuras</b>	<b>xvii</b>
<b>Lista de Tabelas</b>	<b>xix</b>
<b>Nomenclatura</b>	<b>1</b>
<b>1 Introdução</b>	<b>3</b>
1.1 Motivação . . . . .	5
1.2 Objetivos . . . . .	6
1.2.1 Geral . . . . .	6
1.2.2 Específicos . . . . .	6
1.3 Estrutura do Texto . . . . .	7
<b>2 Revisão Bibliográfica</b>	<b>9</b>
2.1 Considerações Iniciais . . . . .	9
2.2 Holística . . . . .	10
2.2.1 Abordagem Estatística . . . . .	10
2.2.2 Abordagem Baseada em Aprendizado de Máquina . . . . .	15
2.2.3 Vantagens e desvantagens . . . . .	16
2.3 Extrator de característica . . . . .	17

2.3.1	Vantagens e desvantagens . . . . .	19
2.4	Formas de lidar com a variação de iluminação . . . . .	20
2.4.1	Abordagem passiva . . . . .	20
2.4.2	Abordagem ativa . . . . .	23
2.5	Considerações Finais . . . . .	25
<b>3</b>	<b>Marco teórico</b>	<b>27</b>
3.1	Abordagem Holística . . . . .	28
3.1.1	<i>Eigenfaces</i> . . . . .	28
3.1.2	<i>Fisherface</i> . . . . .	31
3.2	Abordagem Extração de Características . . . . .	34
3.2.1	SIFT . . . . .	34
3.2.2	Mel e Mellin-cepstrum . . . . .	40
3.2.3	Transformada Discreta de Cosseno - DCT . . . . .	45
3.3	Classificação . . . . .	49
3.3.1	SVM . . . . .	49
3.4	Considerações finais . . . . .	52
<b>4</b>	<b>Modelo Proposto</b>	<b>55</b>
4.1	Considerações iniciais . . . . .	55
4.2	Normalização da iluminação . . . . .	56
4.2.1	No domínio espacial (Realce do Contraste Local - LCE) . . . . .	56
4.2.2	Normalização da iluminação no domínio DCT - Transformada Logarítmica . . . . .	59
4.3	Extração de características . . . . .	64
4.4	Redução de dimensionalidade . . . . .	65

4.5	Classificação . . . . .	69
4.6	Considerações finais . . . . .	69
<b>5</b>	<b>Resultados</b>	<b>71</b>
5.1	Descrição dos bancos de dados . . . . .	71
5.2	Protocolo de experimentos . . . . .	74
5.3	Experimento I . . . . .	75
5.4	Experimento II . . . . .	77
5.5	Experimento III . . . . .	77
5.6	Considerações Finais . . . . .	82
<b>6</b>	<b>Conclusão</b>	<b>85</b>
	<b>Referências Bibliográficas</b>	<b>89</b>





# Lista de Figuras

3.1	Modelo de Reconhecimento de Face . . . . .	27
3.2	A mesma pessoa com diferentes condições de iluminação (Belhumeur, Hespanha and Kriegman 1997b) . . . . .	29
3.3	Reta indicando a posição de maior separabilidade dos conjuntos (Santos 2005). . . . .	32
3.4	Comparação entre o Eigenface e o Fisherface para um problema de duas classes (Santos 2005). . . . .	35
3.5	Visualização da Diferença do Conceito de Significado (Lowe 2004). . . . .	36
3.6	Comparação de pontos com 26 vizinhos com uma matriz $3 \times 3$ (Lowe 2004). . . . .	37
3.7	Gradientes da imagem e descritor dos pontos de interesse, respectivamente (Lowe 2004). . . . .	39
3.8	(a) Magnitude do <i>Mel-cepstrum</i> 2D $35 \times 35$ de uma face e (b) <i>Mellin-cepstrum</i> 2D $35 \times 35$ da matrix da imagem de face (Cakir and Cetin 2011). . . . .	41
3.9	Representação da grade 2D <i>Mel-cepstrum</i> no domínio DTFT (Cakir and Cetin 2011). . . . .	41
3.10	Diagrama da técnica <i>Mel-cepstrum</i> (Cakir and Cetin 2011). . . . .	42
3.11	Diagrama da técnica <i>Mellin-cepstrum</i> (Cakir and Cetin 2011). . . . .	43
3.12	$M \times M$ com pesos normalizados para enfatizar a alta frequência (Cakir and Cetin 2011). . . . .	44
3.13	Sistema de reconhecimento de face proposto por (Kao, Hsu and Yang 2010). . . . .	46

3.14	Hiperplano separador das classes representado pela linha cheia. O tamanho da margem é indicado por $\gamma$ (Batista 2009). . . . .	51
3.15	(a) Conjunto de dados não-linear. (b) Fronteira curva no espaço de entradas para a separação das classes. (c) Fronteira linear no espaço de características (Batista 2009). . . . .	51
4.1	Modelo Proposto de Reconhecimento de Faces Invariante a Iluminação. . . . .	56
4.2	Reiluminação de faces (Wang, Liu, Hua, Wen, Zhang and Samaras 2007). . . . .	57
4.3	Antes e depois da aplicação do LCE (Kao, Hsu and Yang 2010) . . . . .	59
4.4	Exemplo de compensação da iluminação no domínio DCT. . . . .	61
4.5	Desvio padrão dos coeficientes DCT (Chen, Meng and Shingian 2006). . . . .	63
4.6	Descartando coeficientes DCT (Chen, Meng and Shingian 2006). . . . .	63
4.7	(a) Abordagem determinística Ziguezague; (b) Abordagem determinística Máscara Zonal (Dabbaghchian, Ghaemmaghami and Aghagolzadeh 2010) . . . . .	65
5.1	Exemplo de imagens das bases de dados (a) PIE e (b) Yale. . . . .	72
5.2	Exemplo de imagens das bases de dados (a) JAFFE e (b) AT&T. . . . .	73
5.3	Exemplo de imagens das bases de dados (a) Georgia e (b) UMIST. . . . .	73
5.4	Gráfico comparativo para a banco de dados YaleB. . . . .	79
5.5	Gráfico comparativo para a banco de dados PIE. . . . .	79
5.6	Resultado comparativo para a banco de dados JAFFE . . . . .	82
5.7	Resultado comparativo para a banco de dados AT&T . . . . .	83
5.8	Resultado comparativo para a banco de dados UMIST . . . . .	83
5.9	Resultado comparativo para a banco de dados Georgia . . . . .	84

# Lista de Tabelas

3.1	Principais <i>kernels</i> utilizado no SVM (Lorena and Carvalho 2007). . . . .	52
5.1	Características dos bancos de dados utilizados . . . . .	74
5.2	Variações do modelo proposto no presente trabalho . . . . .	76
5.3	Resultados das variações do modelo proposto LCE+DCT, LCE+DCT+DPA, Normalização no domínio DCT+DCT e Normalização no domínio DCT+DCT+DPA. A variável (Acc) representa a acurácia e a variável (Var) a variância. . .	76
5.4	Resultados das variações do modelo proposto LCE+DCT, LCE+DCT+DPA, Normalização no domínio DCT+DCT e Normalização no domínio DCT+DCT+DPA. A variável (Prec) representa a precisão e a variável (Rev) a revocação. . .	77
5.5	Resultado dos experimentos considerando a acuracia e a variância. A acuracia é representada por (Acc) e a variância por (Var) para as bases de dados YaleB e CMU PIE. . . . .	78
5.6	Resultado dos experimentos em termos da precisão e revocação. A precisão é representada por (Prec) e a revocação por (Rev) para as bases de dados Yale B e CMU PIE. . . . .	78
5.7	Resultados dos métodos Eingenface, Fisherface, SIFT, Mel, Mellin, Modelo Proposto com DPA. A variável P representa a precisão e a variável R a revocação. . . . .	81
5.8	Resultados dos métos Eingenface, Fisherface, SIFT, Mel-cepstrum, Mellin-cepstrum e o Modelo Proposto. A variável <i>Acc</i> representa a acuracia e a variável <i>Var</i> a variância. . . . .	81



# Nomenclatura

AHH	<i>Adaptative Histogram Equalization</i>
CMU	<i>Carnegie Mellon University</i>
DCT	<i>Discrete Cosine Transform</i>
DFT	<i>Discrete Fourier Transform</i>
DP	<i>Discrimination Power</i>
DPA	<i>Discrimination Power Analysis</i>
ERL	<i>Elastic Local Reconstruction</i>
FT	<i>Fourier transform</i>
FMT	<i>Fourier Mellin transform</i>
FERET	<i>Face Recognition Technology</i>
GPS	<i>Generalized Pattern Search</i>
IC-NPA	<i>Independent Component based Neighborhood Preserving Analysis</i>
ICA	<i>Independent Component Analysis</i>
ISOMAP	<i>Isometric Feature Mapping</i>
JAFFE	<i>Japanese Female Facial Expression</i>

HDR	<i>High Dynamic Range</i>
HE	<i>Histogram equalization</i>
HOGs	<i>Histograms of Oriented Gradients</i>
LBP	<i>Local Binary Pattern</i>
K-nn	<i>K-nearest-neighbor</i>
LCE	<i>Local Contrast Enhancement</i>
LDA	<i>Linear Discriminant Analysis</i>
LLE	<i>Locally Linear Embedding</i>
LPP	<i>Locality Preserving Projections</i>
MMDA	<i>Multi-Manifold Discriminant Analysis</i>
NPA	<i>Discrimination Power Analysis</i>
IO	<i>Interest Operator</i>
OTNPE	<i>Orthogonal Tensor Neighborhood Preserving Embedding</i>
ONPDA	<i>Orthogonal Neighborhood Preserving Discriminant Analysis</i>
PCA	<i>Principal Components Analysis</i>
PIE	<i>Pose, Illumination and Expression</i>
PWC	<i>pair-wise coupling</i>
RBF	<i>Radial Basis Function</i>
RNA	<i>Rede Neural Artificial</i>
SIFT	<i>Scale Invariant Feature Transform</i>
SSS	<i>small sample size</i>
SVM	<i>Support Vector Machine</i>
UMIST	<i>University of Manchester Institute of Science and Technology</i>

# Capítulo 1

## Introdução

As exigências sobre os sistemas de segurança têm se intensificado cada vez mais, e estes por sua vez vem buscando na tecnologia o apoio para uma maior eficiência de suas atividades. No cenário atual dos sistemas de segurança, é de grande importância o reconhecimento e a identificação de indivíduos em tempo e espaço reais. Além disso, são inúmeras as aplicações existentes como a função de identificar um criminoso ou um terrorista em um aeroporto, ou um visitante não autorizado em uma escola para alertar as autoridades competentes. Estas atividades são essenciais e muitas vezes inviáveis para ser realizadas por um humano, por exigir muita atenção sem descanso e maior percepção de pequenos detalhes que geralmente passam despercebidos. Devido a isso vem crescendo a relevância dada à pesquisa de sistemas automáticos de identificação de faces capazes de identificar indivíduos nas mais diversas situações, de forma que auxiliem com eficiência e agilidade o trabalho realizado pelas equipes e sistemas de segurança.

Os seres humanos identificam facilmente uma pessoa através do reconhecimento facial. É muito mais fácil o cérebro memorizar um rosto do que um nome e ele é capaz de realizar essa tarefa mesmo em situações adversas tais como variação de iluminação, distorções ou observações parciais. Não somente o método de reconhecimento facial mas também a identificação do estado emocional através da expressão de um rosto são tarefas triviais para o cérebro humano. Porém pouco se sabe como isso é feito, o que torna ainda mais difícil capacitar uma máquina para essa tarefa.

Implementar um algoritmo que seja capaz de identificar uma face e/ou uma expressão de um rosto em tempo real é um processo bastante complexo. Devido a isso, nos últimos 10 anos o reconhecimento de faces tornou-se uma área popular entre as pesquisas de

visão computacional (Hu 2008a) e o seu maior desafio está nas diversas variações de componentes que podem existir, tais como expressões, envelhecimento, barba, condições de iluminação, uso de óculos, maquiagem, chapéus, disfarces, dentre outros fatores, que tornam a identificação uma tarefa não trivial. Portanto, o objetivo das pesquisas atuais é desenvolver sistemas capazes de aprender e se adequar às variações de aparência, que possam identificar um indivíduo entre várias outras pessoas em tempo real. Contudo ainda há um vasto campo de estudos e pesquisas a serem feitas e diferentes técnicas veem sendo apresentadas. Estes estudos têm sido aprofundados e investigados não somente pela complexidade como também por suas diversas aplicações.

São inúmeros os setores que podem utilizar ferramentas de identificação de faces para auxiliar em suas atividades. Entre eles estão sistemas de entrada em portos nos Estados Unidos, que exigem que visitantes forneçam impressões digitais e uma fotografia digital para identificação (Caleanu 2011). Essa mesma segurança pode ser aplicada em escolas, empresas, universidades, instituições ou até mesmo em estádios e ruas para identificar um fugitivo por exemplo. Sistemas *webs* também se interessam por esta área pois é muito mais prático e seguro identificar a face do que digitar uma senha para *sites* de banco, ou até mesmo em *sites* de relacionamentos proibidos para menores, para reconhecimento dos mesmos. Outro setor que pode utilizar o reconhecimento é da educação, para identificar alunos nas atividades realizadas nos ensinamentos à distância. Também pode ser utilizado no controle parlamentar nas ocasiões de votação e nos sistemas eleitorais. Bem como em estabelecimentos prisionais, na busca de crianças desaparecidas em multidões e em sistemas de segurança residencial para emitir um alarme quando a face não é reconhecida entre os autorizados a entrar na casa.

Um dos fatores mais pertinentes que impossibilita a identificação de faces de forma mais eficiente é referente à variação de iluminação. A mesma está relacionada com o fato da face ser tridimensional e não ser possível controlar a iluminação no mundo real. Devido a isso, são geradas sombras que ocultam características relevantes para o reconhecimento. Com isso, inúmeros estudos veem avançando nesta direção na tentativa de obter o modelo ideal para o tratamento da variação de iluminação com o melhor desempenho.

Contudo, este trabalho propõe um modelo de reconhecimento de face que seja invariante à iluminação. Os testes são divididos em três grupos. O primeiro é formado por dois bancos de dados o Yale B (Georghiades, Belhumeur and Kriegman 2001a) e o CMU PIE (MIT-CMU n.d.), ambos com imagens de face com grande variação de iluminação, usadas para avaliar a abordagem proposta. O segundo, composto também pelos ban-



cos de dados o Yale B (Georghiades, Belhumeur and Kriegman 2001a) e o CMU PIE (MIT-CMU n.d.), é avaliado o desempenho da abordagem proposta em relação à outros métodos invariantes à iluminação. E o terceiro, com os bancos de dados *Japanese Female Facial Expression* (JAFFE) (Kamachi, Lyons and Gyoba 1998) ; *AT&T Laboratories Cambridge database of faces* (Cambrige 2002); *Sheffield* (antigamente conhecida como UMIST) (Graham and Allinson 1998) e *Georgia* (*Georgia Tech Face Database* 2007), para avaliar o desempenho da abordagem em outras situações e em relação à outros métodos apresentados na literatura.

## 1.1 Motivação

É cada vez maior o número de estudos e pesquisas voltados ao reconhecimento de faces para atender as exigências do mundo corporativo e suas atividades em relação a identificação de indivíduos em tempo real e em ambientes também do mundo real. O reconhecimento/identificação automática de pessoas a partir de uma imagem de face é uma tarefa de grande relevância no setor de segurança, como também no que diz respeito a maior confiabilidade de diversos sistemas oferecendo uma grande contribuição para uma vasta área de aplicações.

Contudo o desenvolvimento do presente trabalho ocorre devido ao seu vasto campo de aplicações e por ainda existir muitas possibilidades a ser pesquisadas e exploradas no processo de reconhecimento/identificação facial no que diz respeito à variação de iluminação nas imagens. Esse é um dos principais desafios do reconhecimento de face, pois o não controle da iluminação permite que uma mesma pessoa apareça muito diferente nas imagens, o que pode inviabilizar a identificação da mesma.

Existem várias abordagens propostas para o problema de iluminação (Shan, Gao, Cao and Zhao 2003a, Zhao, Su, Wang and Luo 2003, Zhao and Samaras 2003, Basri and Jacobs 2003, Shashua and Riklin-Raviv 2001, Georghiades, Belhumeur and Kriegman 2001b, Savvides and Kumar 2003, Lee, Ho and Kriegman 2005). Estas podem ser classificadas em três categorias: pré-processamento e normalização, extração de características invariantes, e modelagem.

Métodos como equalização de histograma do inglês *histogram equalization (HE)*, correção de gama, transformação logarítmica pertencentes a abordagem pré-processamento e modelagem, são utilizados para a normalização de iluminação (Shan, Gao, Cao and

Zhao 2003a, Savvides and Kumar 2003). Porém, para variação de iluminação não uniforme, os mesmos não são viáveis (Chen, Meng and Shingian 2006). Com isso, ainda são existentes possibilidades de estudo neste campo que serão abordadas neste trabalho.

## 1.2 Objetivos

### 1.2.1 Geral

Um dos grandes desafios da tarefa de reconhecimento de faces é a invariância à iluminação. Muitas vezes a imagem de reconhecimento possui parte do rosto coberto por sombras que dificultam a visualização de características relevantes para o reconhecimento. Esta condição, de visualização parcial da face, faz com que métodos para identificação de faces sejam pouco eficientes.

Sendo assim, neste trabalho, tem-se como objetivo geral apresentar um método para reconhecimento/identificação de faces que seja invariante à iluminação, ou seja, mesmo a imagem estando com partes escurecidas, seja possível realizar o reconhecimento obtendo um bom desempenho.

### 1.2.2 Específicos

A Transformada Discreta de Cosseno (do inglês *Discrete Cosine Transform* - DCT) é um dos métodos poderosos na extração de características (Dabbaghchian, Ghaemmaghmi and Aghagolzadeh 2010). Algumas propriedades especiais do DCT o torna uma transformada poderosa em aplicações de processamento de imagens, incluindo o reconhecimento de face. Além disso é um método simples de calcular, rápido e geralmente alcança bons resultados (Kao, Hsu and Yang 2010, Chen, Meng and Shingian 2006).

Outro método utilizado no processamento de imagens é o método Realce do Contraste Local (do inglês *Local Contrast Enhancement* - LCE). O mesmo é um eficiente modelo utilizado para o melhoramento da visibilidade das texturas finas. Além disso o modelo é particularmente útil para melhorar o contraste dos detalhes da imagem com Alto Intervalo Dinâmico (do inglês *High Dynamic Range* - HDR).

Para a redução de dimensionalidade um método apresentado por (Dabbaghchian, Ghaemmaghmi and Aghagolzadeh 2010) é chamado de Análise do Poder de Discriminação.

minação (do inglês *Discrimination Power Analysis* - DPA). O mesmo seleciona os coeficientes com maior poder discriminativo que conseqüentemente reduz a dimensionalidade.

Com isso o presente trabalho possui como objetivos específicos:

- analisar o uso do DCT com a técnica LCE para o tratamento da normalização da iluminação, como também com o tratamento da normalização da iluminação no domínio do próprio DCT;
- avaliar o método proposto com o uso de redução de dimensionalidade (DPA);
- avaliar o método proposto com bases de dados com variação de iluminação, como também com variação de expressão, rotação e fundo;
- comparar o método proposto com os modelos invariantes à iluminação apresentados por (Kao, Hsu and Yang 2010, Chen, Meng and Shingian 2006);
- comparar o método proposto com outros métodos clássicos apresentados na literatura.

### 1.3 Estrutura do Texto

Neste primeiro capítulo foram apresentadas as considerações iniciais mostrando o contexto da tarefa de reconhecimento de faces bem como os principais desafios encontrados e as diversas aplicações do mundo real. Foram também apresentadas a justificativa e esclarecidos os objetivos gerais e específicos do presente trabalho. A estrutura do texto a seguir está organizada da seguinte forma.

No Capítulo 2, é visto a revisão bibliográfica que discute as principais abordagens e métodos utilizados apresentados na literatura incluindo os que foram selecionados para comparação do modelo proposto. São também apresentadas as vantagens e desvantagens das abordagens discutidas.

No Capítulo 3, é mostrado o marco teórico do reconhecimento de faces, onde é apresentada de forma detalhada as etapas e como é realizada a tarefa de reconhecimento. São também definidos os métodos implementados ou adaptados para comparar com o modelo proposto.

No Capítulo 4, é apresentado o desenvolvimento o modelo proposto. É apontado em quais autores foram baseados e especificado como é realizado o reconhecimento.

No Capítulo 5, são mostradas as bases de dados que foram utilizadas nos testes tanto do modelo proposto quanto dos métodos de comparação. Também são apresentados os resultados dos testes através de tabelas e gráficos para comparação. E é realizado uma discussão sobre os resultados fazendo o levantamento comparativo entre o modelo proposto e os métodos apresentados na literatura.

E por fim, no Capítulo 6 é feita a conclusão do presente trabalho.

# Capítulo 2

## Revisão Bibliográfica

### 2.1 Considerações Iniciais

A biometria para a tarefa de reconhecimento/identificação de humanos se tornou uma técnica emergente que vem ganhando cada vez mais atenção, assim como formas distintas de aplicações. Isso acontece devido à maior exigência nos mais diversos setores que necessitam de uma maior atenção no quesito de segurança com a identificação de pessoas. As únicas técnicas existentes até pouco tempo atrás para identificação de indivíduos tais como senhas, cartões pessoais e chaves de segurança, permitem com facilidade o roubo, furto, extravio, esquecimento entre outras formas de serem burladas. A tecnologia biométrica possui como ferramentas para identificação características físicas dos indivíduos, como voz, impressão digital, palma da mão, retina e face, o que torna a falsificação muito mais complexa.

Assim como as diferentes possíveis aplicações também existem, na tecnologia biométrica, formas diferentes de implementá-las. A tarefa de reconhecimento pode ser dividida em duas abordagens (Jafri and Arabnia 2009): holística e extração de características. A principal diferença entre as duas abordagens é a forma dos dados de entrada do sistema. O método holístico utiliza toda a imagem como entrada, ou seja, considera todas as informações contidas na imagem. Já a abordagem de extração de característica realiza um pré processamento da imagem de face, extraindo os atributos, onde os mesmos são fornecidos ao sistema para o reconhecimento.

Essa diferença entre as duas abordagens também define as situações em que cada método será mais apropriado. Na abordagem holística, utilizar todas as informações da

imagem, faz com que os dados de entrada sejam mais extensos deixando o sistema mais pesado computacionalmente. Portanto, este método é mais adequado quando se tem uma melhor estrutura computacional. Uma desvantagem desta abordagem é que utilizar todas as informações faz com que aumente a possibilidade de confundir características do fundo da imagem com as que realmente pertencem à face do indivíduo. Por isso a abordagem terá melhor acurácia quando o fundo da imagem for estático.

A abordagem de extração de característica, possui como entrada apenas os atributos extraídos da imagem, por isso a mesma é mais adequada para sistemas estruturalmente simples. O fato dos atributos serem selecionados antes de realizar o reconhecimento, faz com que o sistema tente ignorar a maior quantidade possível de informações do fundo da imagem. Esta característica da abordagem faz com que seja eficaz tanto em imagens com fundo estático quanto em fundo dinâmico.

Nesse capítulo serão apresentados as abordagens para reconhecimento de faces: holística e extração de características. Assim como os métodos que permitem contornar o problema de variação de iluminação.

## 2.2 Holística

A abordagem Holística tem como uma de suas principais características, se basear na descrição total da imagem de face. Assim, tem-se como entrada do sistema de reconhecimento, a imagem da face como um todo. Essa metodologia pode ser dividida em dois subgrupos: abordagem estatística e abordagem baseada em inteligência artificial (Jafri and Arabnia 2009).

### 2.2.1 Abordagem Estatística

Na metodologia estatística, o método mais simples da abordagem holística tem a imagem representada por uma matriz 2D e o reconhecimento é feito pela comparação direta da imagem de entrada com as imagens do banco de dados. Com isso a metodologia fica com algumas limitações tais como ter que possuir condições semelhantes de iluminação, escala, posição da face e fundo. Além disso o método se torna caro computacionalmente e pouco eficiente quando as imagens não estão nas mesmas condições das imagens cadastradas no banco ou possuem algum tipo de ruído (Huang 1998).

Um obstáculo enfrentado para os métodos de reconhecimento é a classificação ser feita em um espaço de alta dimensionalidade. Para uma melhor eficiência do sistema no tratamento deste problema, outros esquemas baseados na redução de dimensionalidade têm sido apresentados. O método mais conhecido entre eles, é chamado de Análise de Componentes Principais (do inglês *Principal Components Analysis* - PCA) (Jain and Dubes 1988) (Fukunaga 1990). O primeiro a propor um método para a redução de dimensionalidade da imagem para reconhecimento facial utilizando PCA foi Sirovich and Kirby (1987). Os autores demonstram que a imagem pode ser representada de forma eficiente por autovalores e que pode ser reconstruída a partir de um pequeno conjunto de autovetores. Mais tarde, Turk and Pentland (1991a) e Turk and Pentland (1991c), baseado no trabalho de Sirovich and Kirby (1987) mostraram que projeções ao longo das autoimagens (*eigenpictures*) podem ser utilizadas como atributos de classificação para reconhecer faces, desenvolvendo um sistema que constrói *eigenfaces* que correspondem aos autovetores associados aos autovalores dominantes da matriz de covariância conhecida, de faces. Então o reconhecimento é feito comparando as projeções dos *eigenfaces* da imagem a ser reconhecida com as das imagens de rostos cadastrada no banco de dados.

Um trabalho da literatura que utiliza a redução de dimensionalidade é apresentado por (Hu 2008a) que propõem um método de reconhecimento facial chamado de análise de Componente Independente Baseado na Preservação de Vizinhança (IC-NPA, do inglês *Independent Component based Neighborhood Preserving Analysis*). A proposta combina o método chamado de Análise de Componentes Independentes do inglês *Independent Component Analysis* (ICA) e a Análise de Preservação de Vizinhança (do inglês *Neighborhood Preserving Analysis* - NPA), onde o NPA é realizado no subespaço ICA reduzido que é construído pelos componentes de imagem. Essa combinação consiste em três estágios, o primeiro usando a técnica PCA para projetar o padrão de face de uma imagem em um espaço de menor dimensionalidade. O segundo estágio utiliza o algoritmo ICA para encontrar características de rosto estatisticamente independentes. E o último estágio é alcançado através do NPA que é usado para encontrar a melhor projeção no subespaço reduzido. Os testes do sistema foram realizados em dois banco de dados, FERET do inglês *Face Recognition Technology* (Phillips, Wechsler, Huang and Rauss 1998) e CAS-PEAL (Gao, Cao, Shan, Zhou, Zhang and Zhao 2004) que foram utilizadas também para testes dos métodos *Eigenface* (Turk and Pentland 1991a), ICA (Bartlett, Movellan and Sejnowski 2002) (Liu 2004), LDA-based (Belhumeur, Hespanha and Kriegman 1997b) (Lu, Plataniotis and Venetsanopoulos 2005) e Laplaciano (He, Yan, Hu, Niyogi and Zhang 2005) para comparação.

Outro trabalho apresentado pelo mesmo ator que também faz uso da redução de dimensionalidade é (Hu 2008b), que utiliza o método holístico estatístico para propor um novo algoritmo de análise de novos sub-espacos lineares chamado de *Orthogonal Neighborhood Preserving Discriminant Analysis* (ONPDA) baseado no algoritmo discriminante linear de Fisher (do inglês *Fisher's linear discriminant*) também conhecido como LDA. Dado um conjunto de pontos de dados no espaço, uma matriz de peso é primeiramente construída, que descreve a relação entre os pontos de dados. Em seguida, a matriz dispersa inter-classe e a matriz dispersa intra-classe são definidas de modo que a estrutura da vizinhança seja preservada na dimensionalidade de espaço reduzido. Portanto, o método apresentado combina efetivamente as ideias da Análise Discriminante Linear (LDA do inglês *Linear Discriminant Analysis*) e as Projeções de preservação de Localidade (do inglês *Locality Preserving Projections* - LPP) e mantém o forte poder de discriminação do LDA enquanto preserva a geometria intrínseca da amostra dos dados.

A abordagem do sistema apresentado por (Turk and Pentland 1991a) e (Turk and Pentland 1991c) para a redução de dimensionalidade tem sido bastante explorada e utilizada de várias maneiras (Pentland, Moghaddam and Starner 1994). Entre as diferentes utilizações está o método proposto de múltipla visão que é aplicado quando há variação de pose na imagem. Neste caso o reconhecimento pode ser feito estimando a posição da face em um autoespaço (*eigenspace*) calculando a partir da combinação do número de imagens com o número de visões das mesmas, ou construindo um conjunto de *eigenspace* separado de visões para cada indivíduo. Entre as duas formas de classificação, a última tem apresentado melhores resultados.

Outra abordagem proposta baseada na variação da aparência da face é a autocaracterística (*eigenfeature*), onde a baixa resolução da imagem é tratada enfatizando as características mais salientes da face, como a boca, o nariz e os olhos. Este método tem apresentado resultados um pouco melhores do que a abordagem *Eigenfaces* básica. Embora não implementado, foi sugerido por Turk and Pentland (1991a) que a variação da escala pode ser tratada utilizando autofaces (*eigenfaces*) multi-escala ou por reescalamento da imagem de entrada para vários tamanhos. O PCA apresenta um bom funcionamento quando se tem apenas uma imagem de cada indivíduo mas quando se tem várias imagens de cada pessoa Belhumeur, Hespanha and Kriegman (1997b) afirmam que, quando é escolhida a projeção que maximiza o total de dispersão, o PCA varia de forma indesejada devido a iluminação e expressão facial. Sendo assim é proposto a utilização do LDA onde é maximizado a relação de dispersão inter-classe e intra-classe o que torna o método supostamente melhor do que o PCA. Esta, também chamada



de *Fisherface*, utiliza a projeção do subespaço para evitar a dispersão da matriz intra-classe de se degenerar e obtém melhores resultados quanto à variação de iluminação e expressões faciais. Os autores de Swets and Weng (1996) também apresentam resultados semelhantes tanto para rostos como também com outros objetos. Porém outros estudos como (Martínez and Kak 2001) mostram que se o banco dados de treinamento for relativamente pequeno o PCA tem uma melhor performance se mostrando mais sensível às diferenças do banco de dados de treinamento.

Ambas abordagens *Fisherface* e *Eigenface* assumem a existência de uma projeção ideal para projetar a imagem não sobrepondo regiões distintas reduzindo o espaço de representação da imagem, na qual cada região corresponde a uma característica da face. Porém, imagens de diferentes pessoas podem mapear frequentemente a mesma região e não serem distinguidas. Para solucionar esta questão, Moghaddam and Pentland (1996) propõem uma abordagem que utiliza a diferença aritmética dos valores correspondentes aos *pixels*. Com isso as imagens são diferenciadas em intra-pessoal, que consiste nas imagens de diferenças geradas a partir de duas imagens da mesma pessoa e extra-pessoal que consiste nas imagens de diferença derivadas de duas imagens de pessoas diferentes.

Tanto a abordagem do *Fisherface* quanto a abordagem do *Eigenface* tem sido apresentadas com inúmeras variações que são muito utilizadas desde a sua apresentação. Entre as variações baseadas no PCA inclui análise do espaço multi-linear (Vasilescu and Terzopoulos 2003), PCA simétrico (Yang and Ding 2003), PCA 2D (Yang and Zhang 2004) (Meng and Zhang 2007), *eigenbands* (Cavalcanti and Filho 2003). Já para as variações baseadas no LDA tem se LDA direto (Yu and Yang 2001) (Song, Zhang, Wang, Liu and Tao 2007), LDA direto-ponderado (Zhou and Yang 2004), LDA de espaço nulo (Chen, Liao, Ko, J. and Yu 2000) (Liu, Wang, Li and Tan 2004b), LDA *dual-space* (Wang and Tang 2004), LDA pares (Loog, Duin and Haeb-Umbach 2001), análise discriminante regularizada (Friedman 1989), decomposição do valor singular generalizado (Howland and Park 2004) (Ye, Janardan, Park and Park 2004), *Direct Fractional Step LDA* (Lu, Plataniotis and Venetsanopoulos 2003). Essas variações do método básico apresentam-se de forma mais eficientes.

Como maior desvantagem dos métodos do LDA e PCA tem-se o fato de que os mesmos analisam somente a estrutura euclidiana e não consegue descobrir a estrutura subjacente caso a imagem sobrepõe sobre alguma subvariedade não linear (Jafri and Arabnia 2009). Portanto tem sido apresentado na literatura algumas técnicas não lineares para descobrir as estruturas subvariadas tais como Mapeamento de características Isométricas (do inglês *Isometric Feature Mapping* ISOMAP) (Tenenbaum, Silva and Langford 2000), *Locally*

*Linear Embedding* (LLE) (Roweis and Saul 2000) (K. and T. 2003), *Laplacian Eigenmap* (Belkin and Niyogi 2001), *Locality Preserving Projection* (LPP) (He, Yan, Hu and Zhang 2003), *Embedded manifold* (Yan, Zhang, Hu, Zhang and Cheng 2001), *Nearest Manifold Approach* (Zhang, Li and Wang 2004), *Discriminant Manifold Learning* (Wu, Chan and Wang 2004) e *Laplacianfaces* (He, Yan, Hu, Niyogi and Zhang 2005).

O PCA faz a busca de autovetores que dependem somente das relações pares (*pairwise*) dos *pixels* nas imagens do banco de dados. Entretanto existem outros métodos que são capazes de encontrar vetores dependentes de relações de ordem superior entre os *pixels*, no qual espera-se melhores resultados. O método ICA é uma generalização do PCA muito usado na tarefa de reconhecimento/identificação. O seu principal objetivo é encontrar uma decomposição e representação independente da imagem, ao invés de uma decomposição não correlacionada da imagem. Bartlett, Movellan and Sejnowski (2002) testam a performance do ICA sobre duas diferentes arquiteturas nas quais uma considera a imagem como variável aleatória sendo os *pixels* resultados, e outra considera os *pixels* como variáveis aleatórias e a imagem como resultado. Bartlett, Movellan and Sejnowski (2002) obtiveram melhores representações que o PCA para reconhecer faces com mudanças de expressões. Outro algoritmo classificador utilizou as duas combinações do ICA e obteve o melhor desempenho, assim como os trabalhos de (Draper, Baek, Bartlett and Beveridge 2003) e (Kwak and Pedrycz 2007) que mostram que a técnica tem melhor desempenho que o PCA na maior parte das circunstâncias.

Outros métodos utilizados na tarefa de reconhecimento são os de subespaço. Foon, Jin and Ling (2004) apresentam melhores resultados que a abordagem básica do *eigenface* integrando a transformada de *Wavelet* à matriz de fatorização não negativa (Lee and Seung 1999). Já no trabalho apresentado por Liu, Wang, Li and Tan (2004a) é construído um subespaço intra-classe com a classificação baseada na distância ponderada do subespaço de cada intra-classe. Um estudo comparativo foi feito entre as representações de face pelos subespaços PCA, ICA, *Fisher Discriminant Analysis* (FDA) e o *Eigenface* Probabilístico por (Li, Zhou and Shekhar 2003), assim como é apresentado em (Yang and Tang 2004), os avanços em subespaço.

Outro método utilizado no reconhecimento de faces é baseado na reconstrução local elástica (do inglês *Elastic Local Reconstruction* ELR). Xie and Lam (2008) apresentam um novo algoritmo de reconhecimento de faces baseado em uma imagem de vista frontal considerando o efeito da estrutura da face utilizando a metodologia holística estatística. O algoritmo compara duas imagens considerando uma combinação de sequência de blocos da imagem local. Em seguida, é proposto a reconstrução local elástica (ELR) para

medir a similaridade entre os blocos de imagens a fim de medir a diferença entre as duas imagens. Comparado com outros métodos como LDA o método proposto requer apenas uma imagem por tema para o treinamento, o que o torna mais útil para aplicações reais.

## 2.2.2 Abordagem Baseada em Aprendizado de Máquina

Para realizar o reconhecimento de faces utilizando Inteligência Artificial são utilizados os métodos de Rede Neural Artificial (RNA). Esta abordagem é utilizada por DeMers and Cottrell (1993) que faz uso da rede neural auto-associativa no processo de redução do PCA extraído da imagem, para cinco dimensões. Para a classificação dos resultados os autores fazem uso de uma RNA perceptron multi-camadas.

As RNAs também são empregadas no processo de reconhecimento de imagens. No trabalho proposto por Eleyan and Demirel (2005) é apresentado um método onde são obtidos os vetores de características através da utilização do PCA, em seguida, a classificação é realizada, por uma RNA *feed forward*. Os testes feitos mostram melhores resultados que o método *Eigenface*, no qual a classificação é feita baseada no vizinho mais próximo.

No trabalho de Li and Yin (2005) é introduzido um sistema onde a imagem passa primeiro pela transformada de Wavelet para decomposição em três níveis. Em seguida é aplicado o método *Fisherface* em cada uma das três subimagens de baixa frequência. Por fim, os classificadores individuais são fundidos através da RNA *Radial Basis Function* (RBF).

Em Melin, Felix and Castillo (2005) para cada módulo da RNA é atribuído uma das três regiões da face: olhos, boca e nariz. Uma *Sugeno Integral Fuzzy* (tomada de decisão) é então utilizada para combinar a saída dos três módulos fazendo assim o reconhecimento.

Entre os trabalhos de aprendizado de máquina é apresentado por (Zhang, Huang, Li, Wang and Wu 2004) uma abordagem em que uma função de similaridade aprende a descrever o nível de confiança de duas imagens pertencerem a mesma pessoa. Para isso são selecionados os traços faciais através de histogramas de Padrão Binário Local (do inglês *Local Binary Pattern* LBP) Ojala, Pietikainen and Maenpaa (2002) de subregiões da imagem de face e a distância  $\chi^2$  entre os histogramas LBP correspondentes, são escolhidas como características discriminantes. Para selecionar as características LBP mais eficientes e obter a função de similaridade na forma de combinação linear, é aplicado

o algoritmo de aprendizado *AdaBoost* introduzido por (Freund and Schapire 1997).

Outro método da aprendizagem de máquina é o método um-contratodos (Krebel 1999) que é utilizada para a decomposição do problema de reconhecimento facial multi-classe em uma série de problemas de classificação binária. Para este método é treinado um classificador para cada par de classe, sendo ignorado todas as demais e onde as saídas são então combinadas para formar o resultado. Para os classificadores binários com saídas probabilísticas podem ser usados o acoplamento de pares (do inglês *pair-wise coupling* PWC) (Hastie and Tibshirani 1998) para acoplar os resultados em um conjunto de probabilidades e o exemplo de teste é associado com a classe de maior probabilidade. Esta técnica tem como principal desvantagem que, quando o exemplo de teste não pertence a nenhuma das classes, a saída é sem sentido, o que pode afetar o resultado final. Para tratar este problema Moreira and Mayoraz (1998) apresenta uma nova versão do PWC chamada de PWC-CC onde para cada classificador binário é treinado um outro classificador para todas as outras classes. O PWC-CC obtém melhores resultados mas ainda assim apresenta desvantagens. Com isso Li and Yin (2005) propõem o NPWC-CC que supera o PWC-CC. A extração de característica é feita com o PCA e o (SVM) é usado para classificação binária.

O SVM é considerado o algoritmo mais eficaz para tratar a classificação de padrões (Li, Wang and Qi 2004). Isto é feito, primeiro, mapeando os exemplos de treinamento para um espaço de características de alta dimensionalidade. Em seguida um hiperplano ótimo distingue as diferentes características maximizando a distância entre as classes (Burgess 1998a). Este tem sido usado por vários pesquisadores (Li, Wang and Qi 2004, Dai and Zhou 2003, D'enz, Castrill'on and Hern'andez 2003, Guo, Li and Kapluk 2000, Liang, Gong, Pan, Li and Hu 2005) apresentando bons resultados.

### 2.2.3 Vantagens e desvantagens

A principal vantagem do método de reconhecimento holístico é o aproveitamento total das informações existentes na imagem o que tem como consequência a pior desvantagem, isto é, levar em consideração todos os *pixels* tem um custo computacional bastante significativo como também exige um alto grau de correlação entre as imagens de treino e de teste. A última exigência faz com que a abordagem deixe a desejar no quesito eficiência do reconhecimento, por levar em consideração muitos detalhes como fundo, objetos externos, iluminação e posição da face.

Assim a abordagem vem sendo acompanhada de técnicas de redução e modificações para compensar tais variações para suprir as questões que levam à baixa discriminação e conseqüentemente ao reconhecimento ineficiente. Com isso, métodos baseado na abordagem holística vem ganhando destaque.

## 2.3 Extrator de característica

Nesta abordagem são identificadas e extraídas medidas distintas da face como distância entre olhos, nariz e boca, suas localizações e estatísticas locais, bem como outros pontos relevantes. Assim o reconhecimento é feito sobre essas características o que reduz a imagem de entrada do sistema a um vetor de características. Para realizar a identificação são empregadas técnicas de reconhecimento de padrões que combinam as medições da face a ser identificada com as características das imagens que estão no banco de treinamento.

O primeiro trabalho automatizado de reconhecimento de faces é baseado no método de extração de característica. Este foi realizado por Kanade (1973) que localiza e extrai automaticamente 16 parâmetros faciais tais como cantos dos olhos, nariz e boca utilizando métodos simples de processamento de imagens. A partir desses 16 parâmetros é então construído o vetor de características utilizado como entrada do sistema de reconhecimento. O autor usou a razão das distâncias, áreas, ângulos e a distância Euclidiana como medida para comparar o vetor de teste com os vetores das imagens cadastradas no banco de dados e definir a similaridade entre elas para atribuir ou não a um indivíduo. Mais tarde Brunelli and Poggio (1993) criaram um sistema de reconhecimento que extrai 35 características geométricas da face baseado no modelo de Kanade (1973). Os autores aplicam a técnica na banco de dados utilizada por Kanade (1973) alcançando um melhor desempenho.

As técnicas mais sofisticadas da abordagem de extração de característica envolvem modelos deformáveis (Yuille, Cohen and Hallinan 1998, Roeder and Li 1995, Colombo, Bimbo and Magistris 1995), métodos baseados na transformada de Hough (Nixon 1985), Operador simétrico de Reisfeld (Reisfeld 1994), no filtro de Graf e operadores morfológicos (Graf, Chen, Petajan and Cosatto 1995). Contudo, todas essas técnicas dependem fortemente de heurísticas tais como a restrição do espaço de busca a partir de restrições geométricas. Esses são modelos tais como os baseados em operadores de simetria e operadores morfológicos apresentados por (Reisfeld 1994) e (Graf, Chen, Petajan and Cosatto 1995) respectivamente. No trabalho de Cox, Ghoson and Yianilos (1996) é

alcançado uma performance considerável utilizando um banco de dados de 685 imagens onde foram usadas 35 características faciais extraídas manualmente.

Caleanu (2011) utiliza o método de extração de característica para propor um sistema de reconhecimento facial utilizando a combinação entre a técnica Operador de Interesse (IO do inglês *Interest Operator*) e o classificador K-vizinhos mais próximos (K-NN do inglês *K-nearest-neighbor*) tendo os parâmetros determinados pelo algoritmo de Busca de Padrão (do inglês *Pattern Search*). Foram considerados dois tipos de algoritmo para a busca de padrões, Busca Adaptativa de Malha (MADS do inglês *Mesh Adaptive Search*) e Busca de Padrão Generalizada (GPS do inglês *Generalized Pattern Search*). Os algoritmos considerados diferem na forma como são calculados os pontos de interesse.

Outros trabalhos da literatura baseados na abordagem de extração de características são apresentados a seguir. Tsao, Lee, Lui, Chang and Lin (2010) propõem um método baseado no aprendizado de máquina para detecção de faces. Na fase de treinamento são extraídas as bordas da imagem usando o operador de Sobel. Logo o algoritmo MAFIA (Burdick, Calimlim and Gehrke 2001) é utilizado para encontrar os padrões mais frequentes nas bordas (características positivas e negativas). Na construção do detetor de faces, são utilizados três classificadores em cascata: classificador de variância, classificador de características da face e classificador *kdtree-based support vector machine*. Yang, Sun and Zhang (2011) propõem um novo método chamado *Multi-Manifold Discriminant Analysis* (MMDA) para extração de característica das imagens e reconhecimento de padrões baseado no aprendizado incorporado em um grafo e o modelo de análise discriminante do Fisher. Em um MMA, os grafos intra-classe e inter-classe, são respectivamente utilizados para caracterizar a compactação intra-classe e a separabilidade inter-classe. O objetivo do MMDA é minimizar a distância intra-classe e maximizar a distância entre as classes. Liu and Ruan (2011) propõem um novo modelo de algoritmo chamado *Orthogonal Tensor Neighborhood Preserving Embedding* (OTNPE) utilizado na redução de dimensionalidade do vetor de característica extraído da imagem.

Além do reconhecimento de faces os métodos de extração de característica também são utilizados para outros tipos de identificação como de expressões faciais proposto por (Kyperountas, Tefas and Pitas 2010). Os autores apresentam um novo método para classificação de expressões. O processo de classificação é dividido em múltiplos problemas de duas classes. Para cada problema, um processo de seleção de características que utiliza uma medida de separação de classe, é utilizado para selecionar as características mais relevantes.

Entre os exemplos de métodos que utilizam essa abordagem está a Transformação de Características Invariantes à Escala (do inglês *Scale Invariant Feature Transform* SIFT) é um descritor local que foi desenvolvido por (Lowe 2004) onde o seu objetivo é descrever os pontos de interesse de uma imagem através da extração de características, sendo invariante às mudanças de escala, luminosidade, ruído, rotação e translação (Lowe 2004). O mesmo está entre os métodos mais utilizados em abordagens de histogramas de palavras visuais, além disso está também entre os que obtém melhores resultados em diversas aplicações (Nowak, Jurie and Triggs 2006, Jiang, Ngo and Yang 2007).

Outro método baseado na extração de características é a análise *Mel-cepstrum*. A mesma é uma das técnicas mais populares de extração de características para aplicação de reconhecimento de voz. E o cepstrum 2D é usado na literatura para detecção de sombra, remoção de eco, controle automático de intensidade, melhoramento de recursos repetitivos e filtragem cepstral (Toreyin and Cetin 2009, Yeshurun and Schawartz n.d., Lee, Kabrisky, Oxley, Rogers and Ruck 1993). Porém no trabalho de Cakir and Cetin (2011) são propostos o *Mel-Cepstrum* e *Mellin-cepstrum* 2D para aplicação em reconhecimento de faces, onde são usados para representar imagens ou regiões da imagem.

Um dos métodos mais importantes da abordagem de extração de características é o DCT (Chen, Meng and Shingian 2006). Algumas propriedades especiais do mesmo, o torna uma transformação poderosa em aplicações de processamento de imagens, incluindo o reconhecimento de face. O DCT foi utilizado em trabalhos como os apresentados por (Kao, Hsu and Yang 2010, Chen, Meng and Shingian 2006)

### 2.3.1 Vantagens e desvantagens

As vantagens existentes nos métodos de extração de característica estão na representação compacta da imagem de face por um vetor de características o que permite melhor desempenho, e por ser um método relativamente robusto no tratamento de variações de posição na imagem de entrada.

Em contra partida está a desvantagem de ser uma abordagem com dificuldades de detectar características automaticamente, identificar as características mais relevantes e não ter nenhum processo que compense este ponto desfavorável.

## 2.4 Formas de lidar com a variação de iluminação

Os métodos de tratamento de variação de iluminação podem ser divididos em duas abordagens: abordagem passiva e abordagem ativa. Na abordagem passiva o método tenta superar o problema de variação da iluminação, estudando o espectro visível da imagem onde a aparência facial foi alterada pela iluminação não uniforme. Já a abordagem ativa, o problema é tratado pelo uso de técnicas para obter imagens de rostos capturadas em condições de iluminação consistentes ou imagens de modalidades invariantes à iluminação (Zou, Kittler and Messer 2007).

### 2.4.1 Abordagem passiva

A abordagem passiva pode ser dividida em quatro grupos: modelo de variação de iluminação, características invariantes a iluminação, normalização fotométrica e modelo *morphable* 3D.

#### Modelo de variação de iluminação

O modelo de variação de iluminação pode ser baseado em um modelo estatístico ou modelo físico. Para o modelo estatístico não é necessário nenhuma suposição sobre a propriedade de superfície e técnicas de análise estatísticas, tais como Eigenface e Fisherface, são aplicadas no conjunto de imagens de treino para alcançar um sub-espço para cobrir a variação de iluminação. E no modelo físico, o processo de formação da imagem é baseado na suposição de determinados objetos pertencentes à reflectância da superfície da imagem que é uma característica invariante à iluminação, tais como reflectância lambertiana, ou seja, que não absorve luz.

#### 1. Subespaços lineares

Hallinan (1994) apresentou um modelo em que foram usados cinco eigenfaces para representar as imagens de faces em uma série de condições de iluminação. Shashua (1997) propôs um método *Photometric Alignment* para encontrar conexões algébricas de um objeto, entre todas as imagens sobre diversas condições de iluminação. Uma ordem  $k$  do modelo de reflectância linear, para qualquer ponto  $p$  da superfície da imagem é definido pelo produto escalar  $x \cdot a$  onde  $x$  é um vetor de tamanho  $k$  de propriedades de superfícies invariantes (superfície normal, sem



alterações de iluminação) e  $a$  é um vetor arbitrário. A intensidade da imagem, representada por  $Im(p)$ , de um objeto com modelo de reflexão de ordem  $k$  pode ser incubido por uma combinação linear de um conjunto de imagens  $k$  do objeto. Para a superfície lambertiana sob fontes pontuais e na ausência de sombras, todas as imagens estão em um subespaço linear 3D de alta dimensionalidade e podem ser representadas por um conjunto de três imagens desta superfície com menos de três fontes de luz conhecidos e linearmente independentes.

Belhumeur, Hespanha and Kriegman (1997a) propôs um método chamado de subespaço linear 3D para o reconhecimento de faces invariante à iluminação que é uma variante do método alinhamento fotométrico. Neste modelo, para gerar a base do subespaço linear 3D, são utilizadas três ou mais imagens de uma mesma pessoa sobre iluminações diferentes. O reconhecimento é feito por comparação da distância entre a imagem de teste e cada subespaço linear. O método também utiliza o Fisherface para maximizar a razão entre a dispersão inter classes e dentro da classe para um melhor desempenho. A superfície normal da imagem escurecida e esbranquiçada podem ser recuperadas.

## 2. Cone iluminação

Belhumeur and Kriegman (1998) mostrou que é possível formar um cone de iluminação convexo a partir de imagens de um objeto convexo do ponto de vista de uma superfície lambertiana mesmo sendo iluminado por um número arbitrário de fontes pontuais. A dimensão do cone de iluminação é o número de superfícies normais distintas e o mesmo pode ser gerado a partir de três imagens em condições distintas de variações de iluminação. Este cone é uma combinação convexa dos raios dados por  $x_{ij} = \max(B_{S_{ij}}, 0)$  onde  $S_{ij} = b_i \times b_j$  e  $b_i, b_j$  são duas diferentes linhas da matriz  $B$  onde cada linha é o produto da superfície esbranquiçada com o vetor de superfície normal.

## Características invariantes à iluminação

Um estudo apresentado por Adini, Moses and Ullman (n.d.) avalia a sensibilidade de várias representações de imagens insensíveis à variação de iluminação. Estas representações incluem mapa de borda, intensidade derivada da imagem e imagem convolucionada com um filtro de Gabor 2D. As representações citadas são realizadas juntamente com uma função *log* para gerar representações adicionais, porém, os experimentos de reconhecimento de face numa banco de dados com variação de iluminação indicam que

nenhuma das representações por si só são suficientes para tratar o problema de variação de iluminação devido à mudança de direção da iluminação.

### 1. Características derivadas de imagens

O mapa de borda é proposto para o reconhecimento de face por Gao and Leung (2002), onde os *pixels* das bordas são agrupados em linhas de segmentos e uma distância de *Hausdoff* é designada para medir a similaridade entre essas duas linhas. Chen, Belhumeur and Jacobs (2000) provaram que não existem funções discriminativas que são invariantes à iluminação mesmo em objetos com superfície Lambertiana, mas mostraram que a probabilidade de distribuição dos gradientes da imagem é uma função geométrica da superfície e da reflectância, que são as propriedades intrínsecas da face. A direção do gradiente da imagem revela-se insensível às mudanças de iluminação e a performance do reconhecimento usando a direção do gradiente fica perto da abordagem cone de iluminação. Características relativas aos gradientes da imagem é aplicado por ? por um robusto sistema de reconhecimento de faces invariante à iluminação. O gradiente da imagem é definido por  $\bar{G}(x, y) = \frac{|\nabla I m(x, y)|}{\max_{(u, v) \in W(x, y)} |\nabla I(u, v)| + c}$  onde  $I m(x, y)$  é a intensidade da imagem,  $\nabla$  é o gradiente operador,  $W(x, y)$  é uma janela local centralizada em  $(x, y)$  e  $c$  é uma constante para evitar divisões por zero.

### 2. Abordagem *Retinex*

Na abordagem *Retinex* a iluminação é estimada pela imagem suavizada que pode ser dividida pela luminância para obter a reflectância. Uma função Gaussiana é aplicada para suavizar a imagem em uma única escala da abordagem *retinex*. Então é aplicado a soma de várias funções Gaussiana com diferentes escalas. A transformada logarítmica é empregada para comprimir a faixa dinâmica nos trabalhos apresentados por ?.

### Normalização fotométrica

Um dos métodos mais utilizados é a equalização de histograma (Gonzalez and Woods 1992). Ao aplicar a equalização no histograma de intensidade dos *pixels*, o mesmo, na imagem resultante, é plano. Uma questão que vale ressaltar é que mesmo em imagens com iluminação controlada, a utilização da técnica também resulta em um melhor desempenho do reconhecimento de faces (Short 2006).

Shan, Gao, Cao and Zhao (2003b) propuseram uma correção de intensidade gamma para a normalização de iluminação. A imagem  $G(x, y)$  corrigida pode ser obtida através da realização de um mapeamento de intensidade  $G(x, y) = cIm(x, y)^{\frac{1}{\gamma}}$ , onde  $c$  é um parâmetro de esticamento, e  $\gamma$  é o coeficiente gamma.

### Modelo morphable 3D

Um modelo de reconhecimento de face baseado no modelo *morphable* 3D foi proposto por Blanz and Vetter (2003). O mesmo descreve a forma e a textura da face separadamente com base na análise do PCA. Para ajustar uma imagem de posição e iluminação desconhecida ao modelo, é necessário um pré-processamento para aperfeiçoar os coeficientes de forma e textura junto com os vinte e dois parâmetros de renderização para minimizar a diferença da imagem de entrada com a imagem renderizada com esses coeficientes. Os parâmetros de renderização incluem intensidade de luz ambiente, rotação, intensidade dirigida da luz, cor entre outros. O modelo apresentado por Phong é baseado no processo que descreve a reflexão difusa e especular da superfície. Primeiramente as imagens de teste e de treinamento são ajustadas para o modelo, depois o reconhecimento pode ser feito com base nos coeficientes de forma e textura.

### 2.4.2 Abordagem ativa

A abordagem ativa possui esse nome por ter a necessidade de que dispositivos adicionais tais como filtros ópticos, sensores específicos ou fontes de iluminação, sejam envolvidos ativamente para obter diferentes modalidades de imagens de face insensíveis a variação de iluminação. Esse recurso pode ser aplicado em imagens 3D (Bowyer, Chang and Flynn 2004) e em imagens de infravermelho.

### Informações 3D

A informação 3D é uma propriedade intrínseca da face invariante à mudança de iluminação. As informações de superfície normal são utilizadas em algumas abordagens passivas, porém elas são recuperadas da imagem capturada pela câmera de luz visível.

A informação 3D possui várias maneiras de ser representada, entre elas, a mais utilizada é a imagem de profundidade, perfil e curvatura da superfície. Pesquisas sobre abor-

dagens para imagens 3D estão disponíveis em (Bowyer, Chang and Flynn 2004, Bowyer, Chang and Flynn 2006). Para atingir uma melhor performance, pode ser utilizado a modalidade 3D com a modalidade 2D (Bowyer, Chang and Flynn 2004, K. C. Chang, Bowyer and Flynn 2005), porém, é necessário ressaltar que as imagens 2D são capturadas em ambientes controlados e não é claro o quanto que a utilização das duas modalidades contribuem para ambientes não controlados. No trabalho de Kittler, Hilton, Hamouz and Illingworth (2005) foi analisado o espectro completo do processamento de faces 3D a partir de sensores de reconhecimento, os mesmos fazem a cobertura completa da face para representá-las de várias maneiras para o uso do reconhecimento de face. Além disso, o autor faz uma discussão sobre o uso das modalidades 2D e 3D juntas e separadas.

## Infravermelho

As faixas de luzes visíveis variam no espectro eletromagnético de  $0,4\mu m - 0,7\mu m$ . Já o infravermelho varia de  $0,7\mu m - 10mm$  e pode ser dividida em 5 bandas chamadas de *Near-Infrared* (Near-IR) ( $0,7 - 0,9\mu m$ ), o infravermelho de ondas curtas *Short-Wave Infrared* (SWIR) ( $0,9 - 2,4, \mu m$ ), o *Mid-Infrared Wave* (MWIR) ( $3,0 - 8,0\mu m$ ), o infravermelho de ondas longas *Long-Wave Infrared* (LWIR) ( $8,0 - 14,0\mu m$ ), e *Far-Infrared* (FIR) ( $14\mu m - 10mm$ ). O infravermelho possui funcionamento semelhante ao espectro visível, pois contém informações sobre a energia refletida da superfície do objeto. Já o infravermelho térmico está relacionado com a radiação do objeto, na qual depende da temperatura e a emissividade do material do objeto (Kong, Heo, Abidi, Paik and Abidi 2004).

Kong, Heo, Abidi, Paik and Abidi (2004) apresentam uma pesquisa sobre o reconhecimento facial utilizando o infravermelho. Em Wilder, Phillips, Jiang and Wiener (1996) é apresentado fazendo pequenas mudanças de iluminação e para pessoas que não utilizam óculos o uso do infravermelho térmico não é significativo quando comparado com as imagens visíveis. Por outro lado, quando há uma vasta variação de iluminação e expressões faciais, a radiometria térmica alcança melhores resultados (Socolinsky, Selinger and Neuheise 2003, Socolinsky and Selinger 2002). Chen, Flynn and Bowyer (2003) apresentam que o uso de imagens térmicas degrade significativamente mais do que as imagens visíveis no quesito tempo entre a aquisição das imagens. Socolinsky and Selinger (2004) reproduziram o trabalho apresentado por (Chen, Flynn and Bowyer 2003) e mostraram também que, em sistemas de reconhecimento de face mais sofisticados, a diferença do desempenho do algoritmo baseado em imagens térmicas e visíveis é pequena.

A banda de iluminação nas proximidades do infravermelho está na partição refletora do espectro entre as bandas de luz visível e infravermelha. O mesmo possui vantagens tanto em relação a luz quanto o infravermelho térmico, pois, como pode ser refletido por objetos ele funciona como uma fonte de iluminação ativa em contraste com o infravermelho térmico. Além disso, penetra facilmente em óculos, ao contrário do infravermelho térmico.

## 2.5 Considerações Finais

Existem duas abordagens para lidar com o reconhecimento de faces. Uma em que são utilizadas todas as características da imagem para a tarefa de reconhecimento e outra onde as características são selecionadas através de um pré-processamento. Neste capítulo foi abordado sobre as duas formas de reconhecimento de faces além das abordagens existentes para lidar com a variação de iluminação no reconhecimento. Para o presente trabalho foi utilizada a abordagem extração de característica. A mesma, por selecionar as características, é mais apropriada para o reconhecimento de face no mundo real já que as imagens contém fundo variável que podem confundir o sistema.

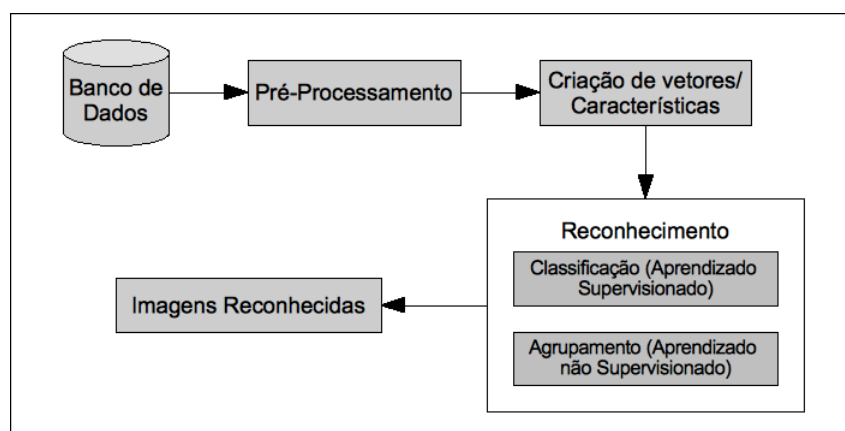
No capítulo a seguir é discutido o passo-a-passo da tarefa de reconhecimento e apresentado métodos de reconhecimento de face da literatura baseados em ambas abordagens: holística e extração de características. Para os métodos invariantes à iluminação são apresentados os métodos utilizados bem como a ferramenta de classificação.



# Capítulo 3

## Marco teórico

O diagrama apresentado na Figura 3.1 demonstra o passo a passo realizado no processo de reconhecimento para as duas abordagens holística e extração de características. Primeiramente o banco de dados é pré-processado, este passo serve tanto para remover ruídos das imagens e outras imperfeições, quanto para tratar o problema de variação de iluminação, ou fazer algum outro tipo de tratamento para alcançar uma melhor performance do método. Depois são aplicadas as técnicas de processamento de imagens para gerar os vetores de características que serão a entrada para o reconhecimento de face. Em seguida é feita a classificação, onde ocorre o reconhecimento propriamente dito, que pode ser uma abordagem baseada no aprendizado de máquina (supervisionada ou não supervisionada) ou outras técnicas de classificação tais como as medidas de similaridade.



**Figura 3.1:** Modelo de Reconhecimento de Face

A diferença entre as duas abordagens está na fase onde são criados os vetores de

características. Na abordagem holística para gerar os vetores são utilizadas todas as informações existentes na imagem, ou seja, toda a imagem é transformada em um vetor de dados. Enquanto que na abordagem extração de características, diferentes atributos são extraídos. Isso acontece devido ao principal desafio enfrentado no processamento de dados multimídia que é a extensão destes dados. Muitos algoritmos e estruturas computacionais se tornam inviáveis diante deste fato. Assim, o desafio está em reduzir a dimensão dos dados sem afetar as informações necessárias.

Portanto, o método extração de características, consiste em extrair os atributos mais relevantes da imagem. Para isso, são utilizados os descritores de imagens que observam características tais como bordas, formas, textura, cores e posições, para formar os vetores de características.

O restante do capítulo discute com mais detalhes o passo a passo da tarefa de reconhecimento e apresenta os métodos clássicos e métodos referentes às abordagens já discutidas no capítulo anterior. Estes foram implementados ou adaptados para comparação com o modelo proposto.

## 3.1 Abordagem Holística

A abordagem holística consiste em empregar todas as características da imagem para o reconhecimento, ou seja, nenhuma informação existente é descartada. Dois clássicos da literatura em processamento de imagens pertencentes a esta abordagem são *Eigenface* e *Fisherface*. Ambos foram adaptados, por já existir diversas implementações na literatura, para teste comparativos com o modelo proposto.

### 3.1.1 Eigenfaces

O PCA, que também é conhecido como expansão de Karhunen-Loève resume-se em reduzir a dimensionalidade dos dados através de uma análise estatística que se baseia na redundância e na variância existentes nos dados. Quando a aplicação da técnica PCA é no reconhecimento de faces, a mesma é chamada de *Eigenface* (Autofaces) (Santos 2005).

O conjunto de Eigenfaces foi originalmente concebido por (Pearson 1901) e desenvolvido independentemente por (Hotelling 1993). Um problema clássico em conjuntos de dados multidimensional é a representação dos dados em um sub-espço de baixa



dimensão. A implementação geral para resolver este problema é representar os dados de modo que a dimensão mais baixa corresponda o tanto possível às dissimilaridades entre os pontos no espaço dimensional original. Um dos mais famosos métodos nessa tarefa é a técnica PCA, que objetiva encontrar um sub-espaço contendo como bases ortonormais, os vetores que definem as principais direções de distribuição dos dados no espaço original otimizando o critério de correlação entre os dados, ou seja, o sub-espaço dos principais componentes minimiza a correlação cruzada entre as amostras de dados (Turk and Pentland 1994). Assim, o método *Eigenface* é baseado na projeção linear da imagem em um espaço de baixa dimensionalidade (Sirovitch and Kirby 1987, Turk and Pentland 1991a, Turk and Pentland 1991b).

Com o uso do PCA para redução de dimensionalidade, o *Eigenface* produz projeções direcionais que maximizam a dispersão entre classe através de todas as imagens de todas as classes. Porém esta técnica retém variações de iluminação indesejados. Na Figura 3.2 é apresentado um exemplo de variação de iluminação. Com isso, enquanto o *Eigenface* é ideal na reconstrução de bases de baixa dimensionalidade, o mesmo pode não ser o ideal do ponto de vista da discriminação.



**Figura 3.2:** A mesma pessoa com diferentes condições de iluminação (Belhumeur, Hespanha and Kriegman 1997b)

Para o reconhecimento de faces utilizando o *Eigenfaces* a partir de uma banco de dados de faces, seja o conjunto de imagens  $\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_M$  que formam o grupo de vetores associados aos *pixels* da imagem  $Im(x, y)$  concatenados em linhas e utilizados no treinamento do sistema, é calculada a média das imagens utilizando a Equação 3.1.

$$\Psi = \frac{1}{M} \cdot \sum_{n=1}^M \Gamma_n \quad (3.1)$$

Em seguida é calculado os desvios de cada imagem com relação a média utilizando a Equação 3.2

$$\phi_i = \Gamma_i - \Psi \quad (3.2)$$

O conjunto de vetores definido por  $\phi$  está normalizado com relação à média e pronto para ser utilizado no cálculo dos autovalores  $\lambda_k$  e autovetores  $\mu_k$ , que forma o conjunto de M vetores ortonormais, da Equação 3.3,

$$\lambda_k = \frac{1}{M} \cdot \sum_{n=1}^M (\mu_k^T \cdot \phi_n)^2 \quad (3.3)$$

onde  $\lambda_k$  é um máximo, sujeito à seguinte Equação 3.4.

$$\mu_l^T \mu_k = \begin{cases} 1, & l = k \\ 0, & \text{caso contrário} \end{cases} \quad (3.4)$$

Os vetores  $\mu_k$  e escalares  $\lambda_k$  são autovetores e autovalores, respectivamente, da matriz de covariância da Equação 3.5,

$$C = \frac{1}{M} \sum_{n=1}^M \phi_n \phi_n^T = A \cdot A^T \quad (3.5)$$

onde a matriz  $A = [\phi_1, \phi_2, \dots, \phi_M]$ .

Assim a matriz  $C$  é de  $N^2 \times N^2$  e determina os  $N^2$  autovetores e autovalores o que é inviável para o tamanho das imagens. Por isso é utilizado um método algébrico que

resolve o problema de  $N^2$ -dimensional. No presente caso resolve os autovetores da matriz  $M \times M$  referente  $A^T \times A$ . Considerando os autovetores  $V_i$  de  $A^T \cdot A$  como a Equação 3.6.

$$A^T \cdot A \cdot V_i = \lambda_i \cdot V_i \quad (3.6)$$

Ao multiplicar ambos os lados por  $A$  obtém-se a Equação 3.7

$$A \cdot A^T A \cdot V_i = \lambda_i \cdot A \cdot V_i \quad (3.7)$$

com isso, os autovetores de  $C = A \cdot A^T$  são  $A \cdot V_i$ .

Assim é gerada a matriz  $H = A^T \cdot A$  (de dimensão  $M \times M$ ), onde  $H_{(m,n)} = \Phi_m^T \cdot \Phi_n$ , e onde estão os  $M$  autovetores de  $V_i$  de  $MAT$  que determinam a combinação linear das  $M$  imagens do grupo de treinamento para formar os autovetores  $\mu_l$  utilizando a Equação 3.8,

$$\mu_l = \sum_{k=1}^M V_l k \phi_k = A \cdot V_l \quad (3.8)$$

onde  $l = 1, 2, \dots, M$ .

Empregando cálculos é reduzido o número de computações de  $N^2$  (número de *pixels* da imagem) para  $M$  (números de imagens). Na prática o conjunto de treinamentos de imagens será relativamente pequeno ( $M \ll N^2$ ).

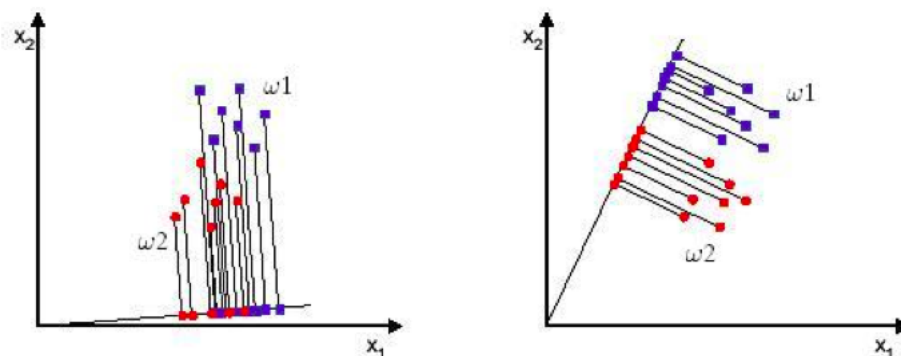
### 3.1.2 Fisherface

A técnica do Fisherface (Turk and Pentland 1991a, Santos 2005) é uma técnica que tem como objetivo maximizar o raio de variância entre as classes e ao mesmo tempo minimizar a variância dentro das classes. Esta tarefa aumenta a separabilidade entre as classes de um conjunto de dados, sendo aplicada uma transformação linear visando encontrar um sistema de coordenadas ótimas para melhor representar os dados com

máxima separabilidade (Santos 2005).

O Fisherface foi utilizado primeiramente em sistemas de reconhecimento de fala para depois ser aplicado nos sistemas de reconhecimento de face no intuito de atingir uma melhor acurácia que o Eigenface. A diferença mais relevante entre as duas técnicas é que o Eigenface proporciona características que capturam as direções principais com relação ao espalhamento dos dados observando as diferenças significativas entre as imagens, porém não reduz o espalhamento das características dentro da classe.

Para entender o funcionamento do Fisherface, considere duas classes distintas em um ambiente 2D. Suponha que um conjunto de amostras igual a  $x^1, x^2, x^3, \dots, x^n$  distribuídas entre duas classes, sendo a classe  $w_1$  com  $N_1$  amostras e a classe  $w_2$  com  $N_2$  amostras. Assim o objetivo é obter uma escalar  $y$  onde as amostras  $x$  são projetadas em uma reta que maximize a separabilidade dos escalares como pode ser observado na Figura 3.3.



**Figura 3.3:** Reta indicando a posição de maior separabilidade dos conjuntos (Santos 2005).

Com isso, busca-se encontrar o melhor vetor para a projeção. O vetor médio de cada classe no espaço original e no espaço Fisherface é obtido pela Equações 3.9 e 3.10

$$\mu_i = \frac{1}{N_i} \cdot \sum_{x \in w_i} x \quad (3.9)$$

$$\tilde{\mu}_i = \frac{1}{N_i} \cdot \sum_{x \in w_i} y = \frac{1}{N_i} \cdot \sum_{x \in w_i} W^T \cdot x = W^T \cdot \mu_i \quad (3.10)$$

O Fisherface propõe maximizar uma função que descreva a diferença entre as médias, normalizada pela medida de dispersão inerente a cada classe. Com isso, é definida para cada classe a medida de dispersão equivalente a variância através da Equação 3.11.

$$\tilde{S}_i^2 = \sum_{y \in w_i} (y - \tilde{\mu}_i)^2 \quad (3.11)$$

Assim, o Fisherface é definido como uma função linear  $W^T \cdot X$  que maximiza a função representados pela Equação 3.12.

$$J(W) = \tilde{\mu}_i^2 = \sum_{y \in w_i} (y - \tilde{\mu}_i)^2 \quad (3.12)$$

onde  $W$  representa o vetor para projeção das amostras onde as características pertencentes a uma mesma classe permanecem muito próximas deixando as médias entre os conjuntos mais distantes.

Como o objetivo é encontrar um vetor  $W$  que defina a melhor projeção, torna-se necessário expressar  $J(W)$  como uma função explícita de  $W$  como na equação 3.13 onde o conjunto de dados com duas classes define  $S_W = S_1 + S_2$  onde  $S_W$  é chamado de matriz de dispersão dentro da classe.

$$S_i = \sum_{x \in w_i} (x - \mu_i) \cdot (x - \mu_i)^T \quad (3.13)$$

Como o método Fisherface é baseado em classe que visa selecionar vetores  $W$  para formar um espaço vetorial onde a dispersão entre as classes e dentro das classes seja maximizada, a matriz de dispersão entre classes é definida pela Equação 3.14

$$S_B = \sum_{i=1}^c N_i \cdot (\mu_i - \mu) \cdot (\mu_i - \mu)^T \quad (3.14)$$

e a matriz de dispersão dentro da classe pela Equação 3.15

$$S_W = \sum_{i=1}^c \sum_{x_k \in X_i} (x_k - \mu_i) \cdot (X_k - \mu_i)^T \quad (3.15)$$

onde  $\mu$  é a média de todo o conjunto,  $\mu_i$  é a média das imagens de classe  $X_i$  e  $N_i$  é o número de amostras na classe  $X_i$ . Caso  $S_W$  seja não-singular, o vetor de projeção ótimo  $W_{otimo}$  é escolhido como a matriz com colunas ortornormais. O mesmo maximiza a razão entre o determinante da matriz de espalhamento inter-classe com relação ao determinante da matriz de espalhamento dentro da classe. Com isso obtém-se a equação 3.16

$$W_{otimo} = argmax \left( \frac{W^T \cdot S_B \cdot W}{W^T \cdot S_W \cdot W} \right) = [w_1, w_2, \dots, w_m] \quad (3.16)$$

onde  $\{W_i | i = 1, 2, 3, \dots, m\}$  é o conjunto generalizado de autovetores de  $S_B$  e  $S_W$  correspondente aos  $m$  maiores autovalores  $\{\lambda | i = 1, 2, \dots, m\}$ .

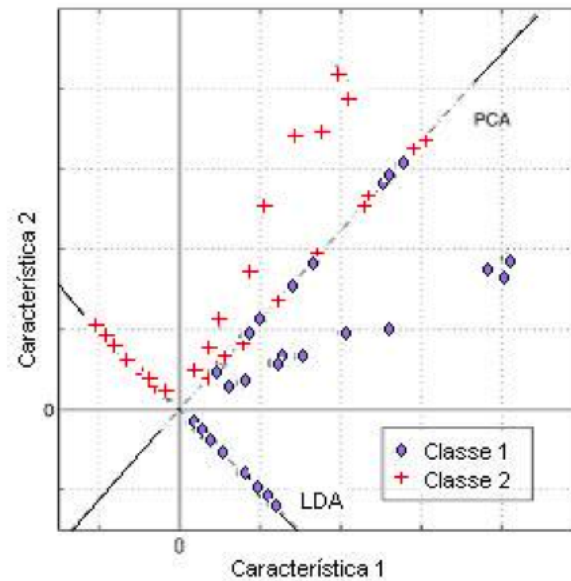
A Figura 3.4 mostra um comparativo do comportamento entre os dois métodos: Eigenface e Fisherface. Como pode ser observado o Eigenface não separa as classes para um sub-espço de projeção 1D como é feito pelo método Fisherface.

## 3.2 Abordagem Extração de Características

Nesta seção são apresentados os métodos SIFT, *Mel-cepstrum*, *Mellin-cepstrum*, Transformada discreta de cossenos apresentado por (Kao, Hsu and Yang 2010) e por (Chen, Meng and Shingian 2006) pertencentes à abordagem extração de características. Estes foram selecionados da literatura recente e foram implementados a partir de artigos publicados.

### 3.2.1 SIFT

O SIFT é um descritor local desenvolvido por Lowe (2004) onde o seu objetivo é descrever os pontos de interesse de uma imagem através da extração de características, sendo invariante às mudanças de escala, luminosidade, ruído, rotação e translação (Lowe



**Figura 3.4:** Comparação entre o Eigenface e o Fisherface para um problema de duas classes (Santos 2005).

2004).

O processo de extração de característica do método SIFT consiste em quatro fases. Na primeira fase é construída uma pirâmide de imagens onde é feita uma busca por características estáveis em várias escalas através da função Gaussiana contínua. Estas são as características locais invariantes a mudança de escala da imagem. Para isso, é definido o espaço da escala pela função descrita na Equação 3.17 que é obtida pela convolução da imagem de entrada  $Im(x, y)$  com a função Gaussiana da escala variável 3.18,

$$L(x, y, \sigma) = G(x, y, \sigma) * Im(x, y) \quad (3.17)$$

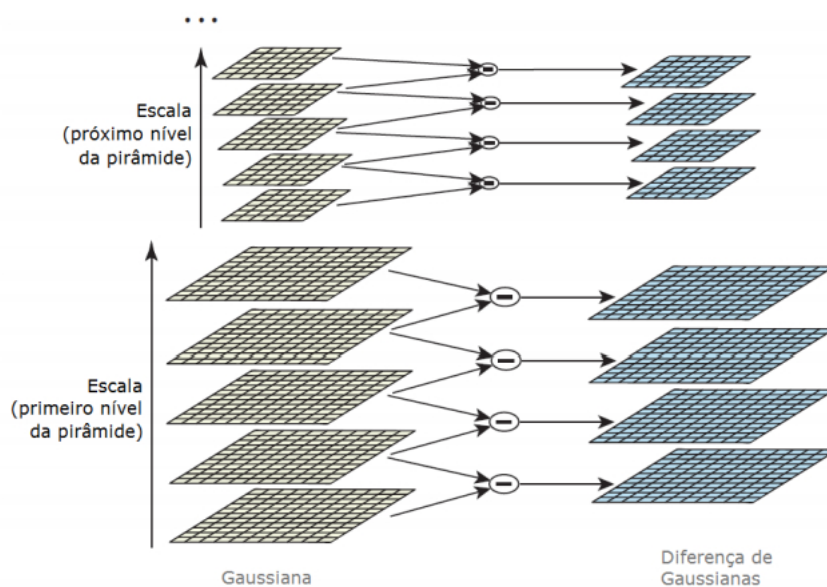
onde \* é a operação de convolução em  $x$  e  $y$ .

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (3.18)$$

Para uma melhor eficiência da detecção dos pontos de interesse estáveis, é então utilizado os extremos no espaço da escala através da convolução da diferença Gaussianas na imagem pela equação 3.19 que é calculada pela diferença entre duas escalas por um valor constante  $k$ .

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * Im(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (3.19)$$

$D(x, y, \sigma)$  é a diferença entre as imagens que são submetidas ao filtro Gaussiano com escalas  $\sigma$  e  $k\sigma$ . Esta diferença de Gaussianos se aproxima ao Laplaciano de uma imagem como discutido em (Lowe 2004). As pirâmides de imagens são utilizadas em imagens de multi-resolução e o Gaussiano de diferentes escalas podem ser construídas usando um tamanho constante do filtro. A pirâmide é ilustrada na Figura 3.5, onde pode ser observado que a imagem inicial é convolucionada de forma incremental para produzir as imagens separadas pelo fator constante  $k$ .

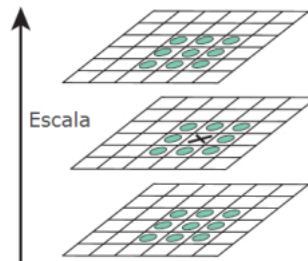


**Figura 3.5:** Visualização da Diferença do Conceito de Significado (Lowe 2004).

Uma vez que a pirâmide foi construída, é então aplicado o detetor de máximos e mínimos locais de  $D(x, y, \sigma)$  para identificar os pontos de interesse na imagem comparando os vinte seis elementos vizinhos em uma matriz  $3 \times 3$ . Como pode ser observado



na Figura 3.6 o *pixel* é comparado com os oito vizinhos mais próximos e com os nove vizinhos abaixo e acima dele.



**Figura 3.6:** Comparação de pontos com 26 vizinhos com uma matriz  $3 \times 3$  (Lowe 2004).

Se o *pixel* for menor ou maior que todos os vinte seis vizinhos comparados, então o mesmo é selecionado como candidato e passa a ser considerado um potencial ponto de interesse.

Na segunda fase, são determinados a localização e a escala de cada ponto de interesse candidato. Nesta fase, também é feita uma seleção de pontos baseada na estabilidade. Os casos limites são ignorados e características como bordas também são eliminadas por ser pontos de baixo contraste e sensíveis a ruídos como definido por (Lowe 2004). Esta tarefa é feita tomando as diferenças das amostras dos pontos vizinhos para calcular a matriz  $2 \times 2$  Hessiana como pode ser observada em 3.20 que calcula a localização e escala dos pontos de interesse candidatos formando a curvatura principal.

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix} \quad (3.20)$$

Os autovalores de  $H$  são proporcionais a curvatura principal de  $D$ . Na abordagem de Harris e Stephens (1988), pode-se evitar calcular os valores próprios e concentrar apenas na sua relação. Sendo  $\alpha$  o valor próprio de maior magnitude e  $\beta$  o de menor magnitude, então a soma dos autovalores da diagonal da matriz de  $H$  e seu produto determinante da Equação 3.21

$$\begin{aligned} Tr(H) &= D_{xx} + D_{yy} = \alpha + \beta, \\ Det(H) &= D_{xx} \cdot D_{yy} - (D_{xy})^2 = \alpha\beta \end{aligned} \quad (3.21)$$

Caso o determinante seja negativo as curvaturas têm sinais diferentes para que o ponto descartado não seja um ponto externo. Agora, sendo  $r$  a razão entre o autovalor de maior e menor magnitude, então  $\alpha = r\beta$ . Em seguida é calculado a Equação 3.22

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha \cdot \beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r + 1)^2}{r} \quad (3.22)$$

que depende somente da razão entre os valores próprios individuais. O valor  $\frac{(r+1)^2}{r}$  é o mínimo quando os dois autovalores são iguais e incrementam com  $r$ . Portanto para verificar se a proporção de curvatura principal está abaixo de um certo valor,  $r$ , basta verificar se é válida a condição da equação 3.23.

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r + 1)^2}{r} \quad (3.23)$$

Na terceira fase, são associados à cada ponto de interesse uma ou mais orientações que se baseiam na direção do gradiente local para obter a invariância referente à rotação. Para isso, é calculado a magnitude e a orientação do gradiente para cada *pixel* utilizando a diferença entre os mesmos.

Então, são gerados histogramas de orientação dos *pixel* da região vizinha do ponto de interesse através da equação 3.24. Os picos do histograma equivalem a direções dominantes dos gradientes locais.

$$h(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2} \quad (3.24)$$

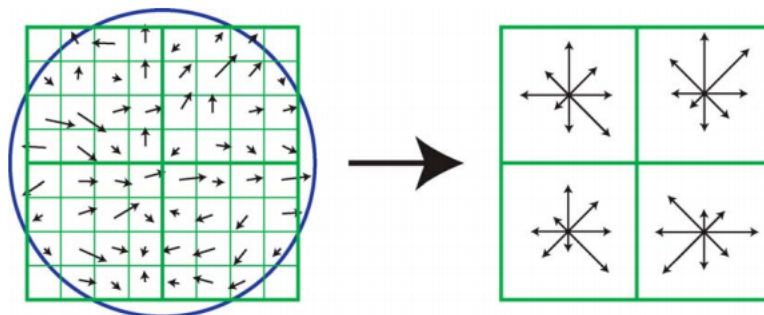
Na última fase, depois dos pontos de interesse serem identificados, os mesmos devem possuir representações distintas quantitativas. Portanto são computados descritores

que demonstram as regiões equivalentes aos pontos de interesse através da magnitude e orientação dos gradientes em volta do ponto de interesse.

Para computar o descritor de interesse, as magnitudes do gradiente e orientação utiliza a escala do ponto para amostrar em volta da localização do ponto de interesse. As setas pequenas de cada localização mostrada no lado esquerdo da Figura 3.7 representam os gradientes pré-computados em todos os níveis da pirâmide. Depois, é utilizada uma função Gaussiana com  $\sigma$  igual à metade da largura da janela do descritor para associar um peso à magnitude do gradiente de cada ponto, onde o objetivo é evitar mudanças repetidas e dar menos relevância aos gradientes distantes, como pode ser observado na Figura 3.7 com a janela circular (Batista 2009).

Em seguida são gerados os histogramas de orientação com as regiões de amostragem e o descritor é formado por um vetor que possui as magnitudes de todas as orientações dos histogramas, nos quais correspondem aos tamanhos das setas apresentadas no lado direito da Figura 3.7. A mesma apresenta uma matriz de histogramas de orientações  $2 \times 2$  que no algoritmo são de dimensões  $4 \times 4$  com um vetor de 128 elementos para cada um dos pontos de interesse que é normalizado.

Assim, cada ponto agora possui posição, escala e orientação e o descritor distingue entre esses pontos. Este, que é um descritor local de amostragem de todos os gradientes em torno de um ponto. Os gradientes são ponderados por uma janela Gaussiana, indicada pelo círculo. As amostras são acumuladas em histogramas de orientações (8 direções) para cada sub-região.



**Figura 3.7:** Gradientes da imagem e descritor dos pontos de interesse, respectivamente (Lowe 2004).

Para o reconhecimento, cada imagem de teste é comparada com o conjunto de imagens de treinamento de forma que todo descritor de característica da imagem é compa-

rado quantitativamente com cada descritor de treinamento. Portanto, quando os vetores estão mais próximos uns dos outros em termos de distância Euclidiana, significa que as características entre as imagens combinam.

### 3.2.2 Mel e Mellin-cepstrum

*Mellin-cepstrum* é um invariante de amplitude e rotação (Cakir and Cetin 2011). Em 2D *Mel-cepstrum* e *Mellin-cepstrum*, a divisão logarítmica da grade da transformada discreta de *Fourier* (do inglês *Discrete Fourier Transform* - DFT) realiza a redução de dimensionalidade. Baixas e altas frequências são combinadas em *bins* de valores frequentes de uma forma logarítmica durante o cálculo do *Mel-cepstrum* 2D.

Para o processo de extração de características o 2D *Mel-cepstrum* utiliza a definição de 2D *cepstrum*, onde  $\hat{I}m(x, y)$  de uma imagem 2D  $Im(x, y)$  é dada pela Equação 3.25,

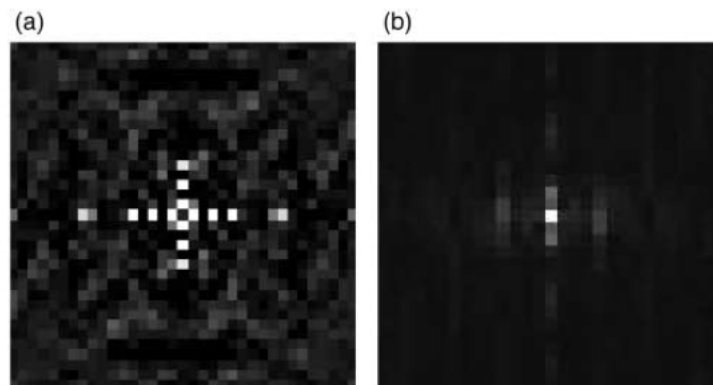
$$\hat{I}m(p, q) = F_2^{-1}(\log|Y(u, v)|^2) \quad (3.25)$$

onde  $(p, q)$  denota coordenadas 2D *cepstrum*,  $F_2^{-1}$  é a transformada inversa de Fourier (IDTFT) e  $Y(u, v)$  é a transformada de Fourier (DTFT) da imagem  $Im(x, y)$ . A sequência *Cepstrum* é de extensão infinita que decai muito rápido (Oppenheim, Schaffer and Buck 1999). Porém nesta implementação os intervalos das coordenadas  $(p, q)$  são os mesmos intervalos da entrada 2D. Na Figura 3.8 são mostrados os coeficientes *Mel* e *Mellin-cepstrum*.

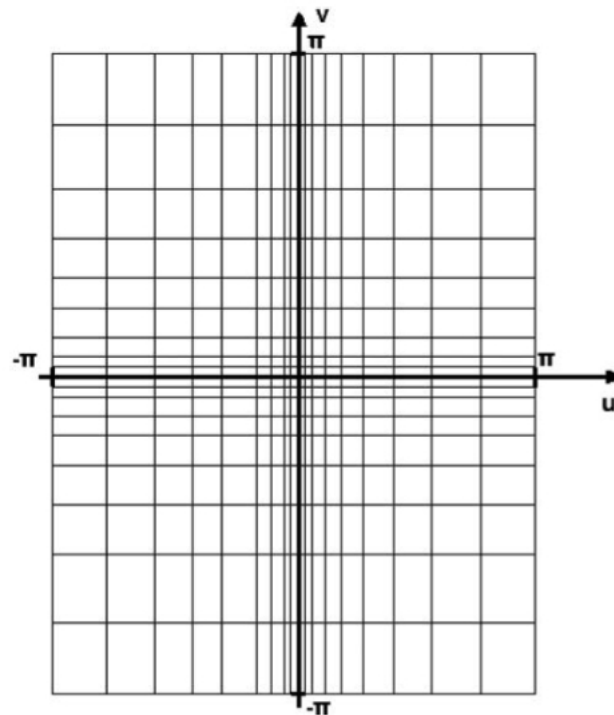
Em 2D mel-cepstrum, os dados de domínio DTFT são divididos em *bins* não uniformes de forma logarítmica como é mostrada na Figura 3.9, onde sua energia  $|G(m, n)|^2$  é calculada pela Equação 3.26,

$$|G(m, n)|^2 = \sum_{k, l \in B(m, n)} |Y(k, l)|^2 \quad (3.26)$$

onde  $Y(k, l)$  é o DFT da imagem de entrada  $Im(n_1, n_2)$ , e  $B(m, n)$  é a grade logarítmica. Os coeficientes de frequência da célula são agrupados para representar a célula correspondente. Em cada malha não uniforme o número e tamanho de células diferem a fim de extrair recursos com características diferentes. Depois desta etapa, os coeficientes



**Figura 3.8:** (a) Magnitude do *Mel-cepstrum* 2D  $35 \times 35$  de uma face e (b) *Mellin-cepstrum* 2D  $35 \times 35$  da matrix da imagem de face (Cakir and Cetin 2011).



**Figura 3.9:** Representação da grade 2D *Mel-cepstrum* no domínio DTFT (Cakir and Cetin 2011).

de frequência  $\hat{I}m(p, q)$  são calculados usando a transformada inversa DFT da Equação 3.27.

$$\hat{I}m(p, q) = F_2^{-1}(\log(|G(m, n)|^2)) \quad (3.27)$$

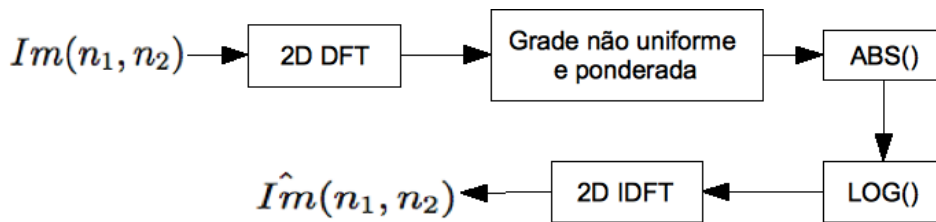
O IDFT é menor que o DFT utilizado para calcular  $Y(k, l)$  devido a grade logarítmica. Uma vez que vários valores DFT são agrupados em cada célula o que resulta na sequência 2D *Mel-cepstrum* usado para calcular o IDFT com menores dimensões que a imagem original.

Para o método 2D *Mel-cepstrum*, primeiro é calculada a  $N \times N$  DFT 2D da imagem de entrada onde  $N$  deve ser maior que a imagem. Durante a computação do DFT para beneficiar com o algoritmo FFT é melhor selecionado um  $N = 2^r$  tal que  $\min_r\{2^r\} > \max\{P, Q\}$  onde  $P$  e  $Q$  são a largura e a altura respectivamente da imagem de entrada.

Depois, a grade não uniforme DTFT é aplicada à matriz resultante DFT e é calculado a energia  $|G(m, n)|^2$  de cada célula. Cada célula da malha pode ser ponderada com um coeficiente e o tamanho dos novos dados é  $M \times M$ , onde  $M \leq N$ .

Em seguida é calculado o logaritmo da energia da célula  $|G(m, n)|^2$ . Por último, o 2D IDFT dos dados de  $M \times M$  é calculada para obter a sequência de  $M \times M$  *Mel-cepstrum*.

A técnica *Mel-cepstrum* pode ser visualizada no diagrama apresentado na Figura 3.10



**Figura 3.10:** Diagrama da técnica *Mel-cepstrum* (Cakir and Cetin 2011).

A técnica de extração de característica 2D *Mellin-cepstrum* é uma versão modificada do 2D *Mel-cepstrum*. Esta técnica utiliza a vantagem da transformada de *Mellin* e fornece características invariantes à rotação, escala e iluminação (Gueham, Bouridane, Crookes and Nibouche 2008). As características de *Fourier-Mellin* são invariantes a escala, rotação e iluminação, porém, com o logaritmo de magnitude do domínio de Fourier, é possível alcançar uma invariância de iluminação no domínio *cepstral*.

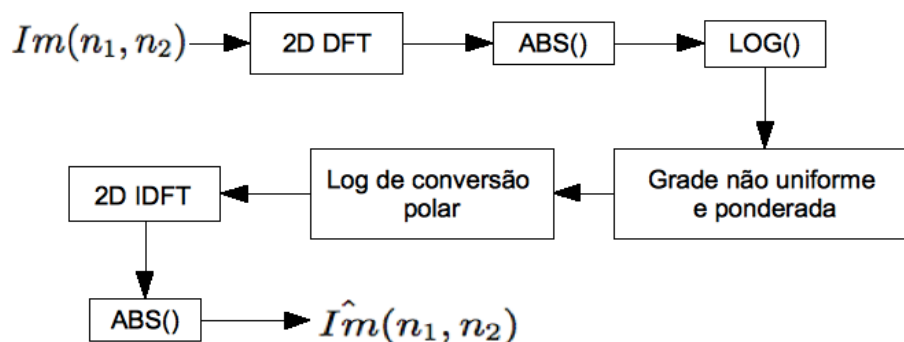
Os passos para a técnica *Mellin-cepstrum* começam também calculando a  $N \times N$

DFT 2D da imagem de entrada onde  $N$  deve ser maior que a imagem, o que facilita a seleção de  $N = 2^r > \text{dimensão}(Im(n_1, n_2))$  para aproveitar o algoritmo FFT durante a computação do DFT.

Depois são calculados os logaritmos da magnitude dos coeficientes da DFT. Em seguida, a malha não uniforme DFT é aplicada à matriz resultante e é calculada a média de cada célula, onde cada célula da malha é representada com esta média e é ponderada com um coeficiente. O novo tamanho dos dados é  $M \times M$  onde  $M \leq N$ .

Então, é feita a conversão do cartesiano para log-polar utilizando interpolação bilinear, onde é atingido a invariância de rotação e escala. Este é um passo fundamental do FMT (do inglês *Fourier Mellin transform*) fornecendo invariância à rotação e escala.

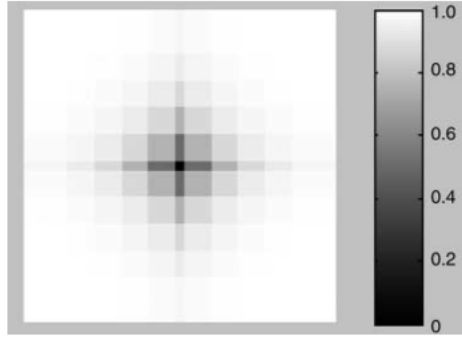
A seguir, é calculado o IDFT 2D dos dados. Finalmente, o valor absoluto ou energia dos coeficientes IDFT são calculados para obter a sequência  $M \times M$  do *Mellin-cepstrum*. O diagrama apresentado na Figura 3.11 apresenta a técnica *Mellin-cepstrum*.



**Figura 3.11:** Diagrama da técnica *Mellin-cepstrum* (Cakir and Cetin 2011).

Em uma imagem, as características mais relevantes são as de alta frequência (bordas, características faciais importantes, formas, boca, nariz, olhos), e para extrair melhores características, os componentes da célula de alta frequência de 2D DFT são multiplicados com maior peso comparado com as baixas frequências. Os componentes de alta frequência são mais enfatizados como resultados. Para isso, os pesos normalizados são organizados como pode ser observado na Figura 3.12, onde os *pixels* brancos correspondem ao valor 1 e os *pixels* pretos ao valor 0. Os pesos são organizados utilizando uma distribuição linear.

Uma característica importante do *cepstrum* é a invariância às mudanças do tamanho do *pixel*. Com isso é possível obter uma boa performance para a invariância à iluminação. Seja  $Y(u, v)$  que denota o DTFT 2D de uma dada imagem  $Im(x, y)$  e  $cIm(x, y)$  tem um



**Figura 3.12:**  $M \times M$  com pesos normalizados para enfatizar a alta frequência (Cakir and Cetin 2011).

$cIm(x, y)$  DTFT para qualquer constante real  $c$ . O *log spectrum* de  $cIm(u, v)$  é obtido através da Equação 3.28 e o *cepstrum* correspondente é dado pela Equação 3.29, onde o  $\delta(p, q)$  corresponde a Equação 3.30.

$$\log(|cIm(u, v)|) = \log(|c|) + \log(|Im(u, v)|) \quad (3.28)$$

$$\psi(p, q) = \hat{a}\delta(p, q) + \hat{Im}(p, q) \quad (3.29)$$

$$\delta(p, q) = \begin{cases} \log 1 & p = q = 0 \\ 0 & \text{caso contrário} \end{cases} \quad (3.30)$$

Portanto, os valores *cepstrum*, com exceção da posição  $(0, 0)$  que corresponde ao termo DC, não variam com as mudanças de amplitude. Uma vez que a magnitude FT (do inglês *Fourier transform*) de  $Im(n_1, n_2)$  e  $Im(n_1 - k_1, n_2 - k_2)$  são as mesmas, o *cepstrum* 2D e *Mel-cepstrum* são características de deslocamento invariante.

Outra característica relevante de 2D *cepstrum* é a simetria:  $\hat{Im}[n_1, n_2] = \hat{Im}[-n_1, -n_2]$ , assim, como resultado, somente metade do 2D *cepstrum* ou coeficientes 2D *Mel-cepstrum*  $M \times M$  são suficientes quando IDFT é utilizado.



### 3.2.3 Transformada Discreta de Cosseno - DCT

A extração de características através do DCT consiste em duas etapas. Na primeira fase são obtidos os coeficientes do DCT e na segunda etapa são selecionados os coeficientes para a construção dos vetores de características. A dimensão da matriz dos coeficientes DCT é a mesma da imagem de entrada. Basicamente o DCT por si só não reduz a dimensionalidade dos dados, devido a isso o mesmo comprime as informações em uma porcentagem de coeficientes.

Para a realização da primeira etapa, em uma imagem 2D onde o tamanho é  $P \times Q$ , os coeficientes DCT são calculados utilizando a equação 3.31,

$$MC(u, v) = \frac{1}{\sqrt{PQ}} \alpha(u) \alpha(v) \sum_{x=0}^{P-1} \sum_{y=0}^{Q-1} Im(x, y) \times \cos \frac{(2x+1)u\pi}{2P} \times \cos \frac{(2y+1)v\pi}{2Q} \quad (3.31)$$

onde  $u = 0, 1, \dots, P, v = 0, 1, \dots, Q$  e  $\alpha(w)$  é definido pela equação 3.32,

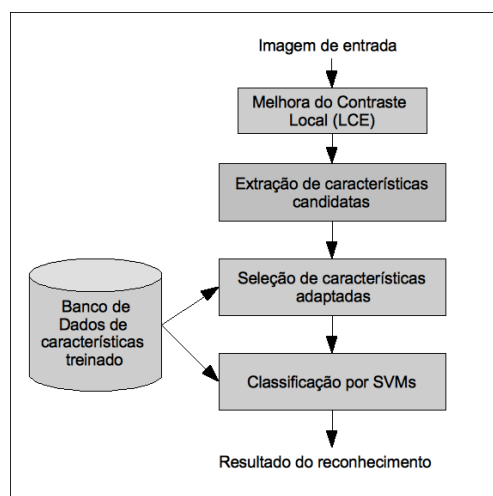
$$\alpha(w) = \begin{cases} \frac{1}{\sqrt{2}} & w = 0 \\ 1 & \text{caso contrário} \end{cases} \quad (3.32)$$

onde  $Im(x, y)$  é a função de intensidade da imagem e  $MC(u, v)$  é uma matriz 2D de coeficientes DCT.

#### Modelo apresentado por Kao

No trabalho apresentado por (Kao, Hsu and Yang 2010) é proposto um modelo de sistema integrado que primeiro compensa iluminação irregular através de um aumento do contraste local. Em seguida as imagens melhoradas são alimentadas em um sistema robusto de reconhecimento facial que seleciona adaptativamente as características mais importantes entre todas as características candidatas e a classificação é realizada por uma máquina de vetores de suporte SVM. A estrutura do modelo é apresentada no diagrama da Figura 3.13

Como pode ser observado na Figura 3.13, a imagem é, antes de tudo, processada pelo método de LCE. Em seguida, a mesma é dividida em blocos de 8 *pixels*. Para ser compatível com os métodos conhecidos de compressão de imagens, todas as características



**Figura 3.13:** Sistema de reconhecimento de face proposto por (Kao, Hsu and Yang 2010).

usadas no reconhecimento são coeficientes do DCT. O classificador SVM é adaptado e mapeia os vetores de características extraídos para um espaço de maior dimensionalidade e encontra os chamados vetores de suporte sobre as bordas de um hiperplano, o que corresponde a um problema de decisão entre duas classes. Uma vez que o SVM só pode determinar um hiperplano de cada vez, uma amostra de entrada desconhecida deve ser processada para cada par de classes arbitrárias. O resultado de classificação final multi-classe é determinada pelo voto dos resultados de classificação, de acordo com todos os hiperplanos. A classe que ganha o número máximo de votos é reconhecido como a saída de reconhecimento final.

Em se tratando da seleção de características para reduzir a dimensionalidade dos vetores, em muitos casos, as imagens disponíveis para o processo de treinamento não é o suficiente. A dimensionalidade da amostragem é geralmente maior que o número de amostras disponíveis para cada classe. Este problema é chamado de Tamanho de Amostra Pequeno (do inglês *small sample size* SSS) e se agrava quando utilizadas ferramentas como PCA e LDA para a reduzir a dimensionalidade.

Um coeficiente DCT pode ser útil para diferenciar duas classes, mas inútil para outras duas classes. Uma opção para esse fato seria incluir mais características da face. Porém, coeficientes DCT também podem ser incluídos baixando ainda mais a taxa de reconhecimento. Com isso, a técnica de análise de característica é a chave para melhorar o sistema de reconhecimento.

Em abordagens tradicionais, tentam extrair valores específicos de características que ajudam à diferenciar faces. Essa redução de dimensionalidade é feita na fase inicial para resolver o problema SSS. No trabalho de Kao, Hsu and Yang (2010) a seleção de características proposta é adaptada utilizando o conceito de hiperplano que é utilizado para definir as classes e é explicado com detalhes na seção 3.3.1. Para um hiperplano  $\Omega_{a,b}$ , somente algumas características são selecionadas, o que acontece da seguinte forma: suponha que o número de amostragem para treinamento seja  $d$  para cada classe. Assim,  $C_a = \{v_{a,1}, v_{a,2}, v_{a,3}, \dots, v_{a,d}\}$  e  $C_b = \{v_{b,1}, v_{b,2}, v_{b,3}, \dots, v_{b,d}\}$  são os conjuntos de vetores de características de treino, onde  $v_{ij}$  é um vetor que inicialmente contém  $h$  valores de características e pode ser representado por  $v_{i,j} = [v_{i,j,1}, v_{i,j,2}, v_{i,j,3}, \dots, v_{i,j,h}]$ . A relação de discriminação  $I_{a,b,k}$  da  $k$ -ésima característica é avaliada pela Equação 3.33 para um hiperplano  $\Omega_{a,b}$ . A Equação 3.33 calcula a separabilidade entre as duas classes  $a$  e  $b$  e a estabilidade da mesma classe para a  $k$ -ésima característica.

$$I_{a,b,k} = \frac{(\mu_{a,k} - \mu_{b,k})^2}{\sigma_{a,k}^2 + \sigma_{b,k}^2} \quad (3.33)$$

onde  $\mu_{i,k}$  e  $\sigma_{i,k}$  denotam a média e o desvio padrão da  $k$ -ésima característica para todas as amostras de treinamento da classe  $I$  e podem ser obtidos pelas equações 3.34 e 3.35 respectivamente.

$$\mu_{i,k} = \frac{1}{d} \sum_{j=1}^d v_{i,j,k} \quad (3.34)$$

$$\sigma_{i,k} = \frac{1}{d-1} \sum_{j=1}^d (v_{i,j,k} - \mu_{i,k})^2 \quad (3.35)$$

Para um hiperplano, as características são avaliadas e organizadas de modo decrescente, ou seja, as melhores características são as localizadas primeiramente no vetor. A seleção dessas características para um hiperplano  $\Omega_{a,b}$  são feitas da seguinte forma: as relações discriminantes de todas as características são primeiramente organizadas em uma matriz  $I_{a,b} = [I_{a,b,1}, I_{a,b,2}, \dots, I_{a,b,h}]$  de forma decrescente. A matriz classificada será denotada como  $I_{a,b}^S = [I_{a,b,1}^S, I_{a,b,2}^S, \dots, I_{a,b,h}^S]$  e o deslocamento correspondente das caracte-

terísticas são denotados como  $A^S = [u_1, u_2, \dots, u_h]$ . Como mencionado anteriormente, muitas vezes a melhor característica utilizada para distinguir duas classes em um hiperplano, pode não servir para outras duas. Com isso, esta técnica é aplicada para cada hiperplano que filtra e classifica as características mais relevantes reduzindo a dimensionalidade e resolvendo o problema do SSS no estágio de extração de característica.

De acordo com os resultados experimentais, apresentado por Kao, Hsu and Yang (2010), usar vetores de características com maior dimensão, nem sempre alcança os melhores resultados. Em alguns casos a taxa de reconhecimento cai drasticamente devido a instabilidade causada quando se tem muitas características. Assim, o modelo proposto por Kao, Hsu and Yang (2010) reduz a dimensionalidade e alcança uma melhor performance de reconhecimento. Mais detalhes do modelo apresentado por Kao, Hsu and Yang (2010) são apresentados no capítulo 4

### Modelo apresentado por Chen

O artigo apresentado por Chen, Meng and Shingian (2006) propõe uma abordagem de normalização de iluminação para remover as variações de iluminação, mantendo as principais características faciais intactas. A ideia principal da abordagem proposta é que as variações de iluminação sejam reduzidas significativamente devido à truncagem dos coeficientes DCT de baixa frequência calculados a partir do modelo logarítmico.

Para isso, o modelo proposto por (Chen, Meng and Shingian 2006) se baseia na abordagem pré-processamento e normalização, onde são aplicadas técnicas para normalizar as imagens e assim serem exibidas de forma estável mesmo com condições de iluminação diferentes. Técnicas tais como, equalização de histograma do inglês *histogram equalization*, correção de gama, transformação logarítmica são utilizados para normalização da iluminação. No entanto, em imagens não uniformemente variadas, é complexo utilizar essas técnicas. Sendo assim, Chen, Meng and Shingian (2006) utilizam a técnica de normalização da iluminação no domínio logarítmico baseada no DCT, *i.e.*, onde é calculado o logaritmo da imagem e posteriormente a transformada DCT. Eliminando os coeficientes DCT de baixa frequência é possível reduzir significativamente a variação de iluminação. Ambas as técnicas são detalhadas no Capítulo 4. Após eliminar os coeficientes de baixa frequência é calculado a inversa DCT. O processo de reconhecimento é executado diretamente na imagem de domínio logaritmo, e a transformada logarítmica inversa é ignorada. Os autores mostraram que o PCA (Eigenfaces) pode ser feito no domínio logarítmico e os mesmos resultados são obtidos como se fosse aplicado no domínio espacial.

Os testes do método foram feitos sobre as bases de dados Yale B e CMU PIE. Segundo os resultados, o modelo proposto melhora o desempenho significativamente para as imagens de face com alta variação de iluminação. Outra vantagem do método é que não é necessário nenhum passo de modelação e pode ser facilmente implementado.

### 3.3 Classificação

A classificação é um processo que pode ser usado para encontrar um modelo que descreva diferentes classes de dados (Elmarsri and Navathe 2005). Em outras palavras, a classificação consiste em analisar uma certa característica dos dados e atribuir a uma classe que foi previamente definida (Berry and Linoff 1997, Han and Kamber 2001).

O processo de classificação consiste em duas fases, a de treinamento e a de teste. Neste caso, são necessário duas bases de dados, para atender ambas as fases do método. Na primeira fase, de treinamento, são definidas as classes existentes na banco de dados com rótulos. Trata-se de um processo de aprendizagem onde os rótulos são criados para indicar a que classe pertence cada vetor de características existentes, e por isso chamado de aprendizado supervisionado (Elmarsri and Navathe 2005).

Na segunda fase, com as classes definidas, é quando acontece a classificação propriamente dita. As características dos dados do segundo banco de dados são examinadas e rotuladas, e atribuídas a classe pertencente de acordo ao processo de aprendizagem realizado.

Neste trabalho foi escolhido para realizar a classificação das imagens as Máquinas SVM. As informações sobre a técnica foram retiradas de Batista (2009)

#### 3.3.1 SVM

O SVM é uma técnica de Aprendizado de Máquina que está sendo cada vez mais utilizada nas mais variadas aplicações de reconhecimento de padrões. A mesma vem apresentando resultados superiores a outras técnicas em diferentes tarefas (categorização de textos, análise de imagens e bioinformática) (Lorena and Carvalho 2007, Jiang, Ngo and Yang 2007, Csurka, Dance, Fan, Willamowski and Bray 2004).

Segundo Lorena and Carvalho (2007), o SVM apresenta como vantagens boa capacidade de generalização, na qual é medida pela sua eficácia na classificação de dados que

não pertence ao conjunto utilizado no treinamento. Além disso apresenta-se robusto com dados de maior dimensão e permite representar espaços abstratos de maneira eficiente devido ao uso das funções *kernels* na não-linearização do SVM, já que o cálculo é mais simples do que a função de mapeamento.

O SVM pertence aos métodos de classificação supervisionada. Sendo assim, dado um conjunto de exemplos rotulados na forma  $(x_i; y_i)$ , onde  $x_i$  denota um exemplo e  $y_i$  representa o seu rótulo, é construído um classificador capaz de prever com exatidão o rótulo de novos dados. O classificador gerado também pode ser interpretado como uma função, a qual recebe um dado  $x$  e fornece uma predição  $y$  (Lorena and Carvalho 2007).

Com isso o objetivo do SVM é a geração de hiperplanos para separar os exemplos positivos e os negativos de cada classe. Isto é feito com a maior margem possível, pois a mesma representa uma medida de confiança da precisão do classificador. Este classificador pode ser da forma linear ou não-linear como é apresentado nas próximas seções.

### SVM linear

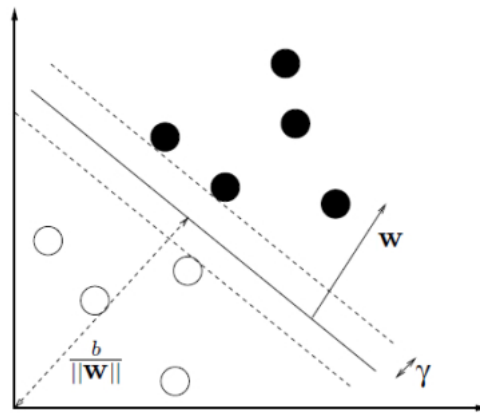
Sendo o conjunto de treinamento representado por  $T$  com  $n$  dados  $x_i \in X$  e os rótulos  $y_i \in Y$ , onde  $X$  compõe o espaço dos dados e  $Y = \{-1, 1\}$ .  $T$  só é separável se for possível separar os dados das classes  $+1$  e  $-1$  através do hiperplano. O mesmo é obtido através da Equação 3.36,

$$f(x) = w \cdot x + b = 0 \quad (3.36)$$

onde  $w \cdot x$  denota o produto escalar dos vetores  $w$  e  $x$ , com  $w \in X$  é o vetor normal ao hiperplano descrito e  $\frac{b}{\|w\|}$  representa a distância entre o hiperplano e a origem, com  $b \in \mathbb{R}$  como mostrado na Figura 3.14

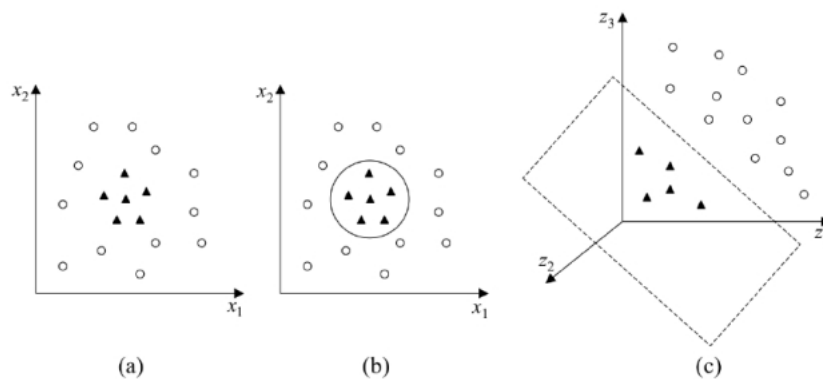
### SVM não linear

Com as diversidades do mundo real, muitas vezes os dados não são linearmente separáveis devido a presença de ruídos e discrepância nos dados. Para este problema o SVM não linear realiza a fase de treinamento mapeando o espaço original para um espaço de maior



**Figura 3.14:** Hiperplano separador das classes representado pela linha cheia. O tamanho da margem é indicado por  $\gamma$  (Batista 2009).

dimensionalidade. Seja  $\Phi : X \rightarrow \mathfrak{Sm}$  o mapeamento em que  $X$  é o espaço original e  $\mathfrak{Sm}$  representa o espaço de características. Na Figura 3.15 pode ser observado em (a, b) exemplos de classes com divisão não linear e em (c) é apresentado um separador SVM linear que pode ser utilizado quando  $\Phi$  é escolhido apropriadamente.



**Figura 3.15:** (a) Conjunto de dados não-linear. (b) Fronteira curva no espaço de entradas para a separação das classes. (c) Fronteira linear no espaço de características (Batista 2009).

Quando utilizada a função  $K$ , nomeada *kernel*, da Equação 3.37 no treinamento, a função  $\Phi$  não necessita ser escolhida de forma explícita se for utilizado uma função  $K$  no treinamento (Burges 1998b),

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (3.37)$$

onde  $x_i$  e  $x_j$  são dois pontos do espaço original e a função  $K$  calcula o produto escalar dos pontos no espaço de características.

Entre os principais *kernels* existentes utilizados no SVM estão o Polinomial, Gaussiano e Sigmoidal. Estes são apresentados na Tabela 3.1.

**Tabela 3.1:** Principais *kernels* utilizado no SVM (Lorena and Carvalho 2007).

Tipo de <i>kernel</i>	Função correspondente	Comentários
Polinomial	$(X_i^T \cdot X_j + 1)^p$	A potência $p$ deve ser especificada pelo usuário
Gaussiano	$\exp(-\frac{1}{2\sigma^2} \ X_i - X_j\ ^2)$	A amplitude $\sigma^2$ é especificada pelo usuário
Sigmoidal	$\tanh(\beta_0 X_i \cdot X_j + \beta_1)$	Utilizada somente para alguns valores de $\beta_0$ e $\beta_1$

O *kernel* mais utilizado nos trabalhos de SVM é o linear ou o Gaussiano que também é conhecido por Função Base Radial (do inglês *Radial Basis Function* RBF) (Jiang, Ngo and Yang 2007, Hsu, Chang and Lin 2009). A escolha do *kernel* juntamente com a escolha dos parâmetros, implica no desempenho do classificador gerado.

Sendo o SVM um classificador binário, para a aplicação em problemas de múltiplas classes, existem abordagens como um-contra-todos (do inglês *one-versus-all*) onde é produzido um classificador para cada classe, cada um separando uma classe de todas as restantes. E todos-contra-todos (do inglês *all-versus-all*) onde são gerados classificadores para separação de cada classe  $i$  de outra  $j$  em que  $i, j = 1, \dots, k$  e  $i \neq j$ .

### 3.4 Considerações finais

Neste capítulo foi discutido o processo de reconhecimento de faces que consiste em três etapas. Na primeira etapa é feito o pré-processamento das imagens para remoção de ruídos e imperfeições. Em seguida são extraídos os vetores de características que podem



ser em toda a imagem quando é baseado na abordagem holística ou selecionados quando é baseado na abordagem extração de características. E por fim é feito o reconhecimento através de classificadores como o SVM.

Entre os métodos de reconhecimento de faces apresentados na literatura, existem dois clássicos: Eigenface e Fisherface que são baseados na abordagem holística e os métodos SIFT, Mel e Mellin-cepstrum que são baseados na abordagem extração de características. Além destes, foram apresentados por Kao, Hsu and Yang (2010) e Chen, Meng and Shingian (2006) dois métodos baseados no DCT também pertencentes a abordagem extração de características, nos quais o método desenvolvido no presente trabalho foi baseado. Ambos os métodos possuem pré-processamento para o tratamento da variação de iluminação como foi discutido neste capítulo.

No próximo capítulo é analisadas em detalhes as técnicas utilizadas para a construção do modelo proposto. Para a normalização da iluminação são apresentadas as técnicas no domínio espacial e no domínio DCT. Em seguida é mostrado o processo de extração de características e a redução de dimensionalidade e por fim a classificação.



# Capítulo 4

## Modelo Proposto

### 4.1 Considerações iniciais

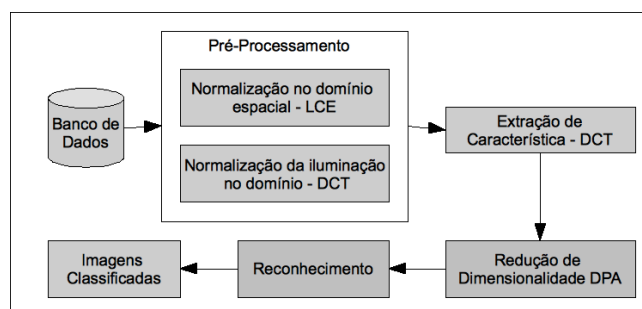
Neste capítulo serão tratados os detalhes do modelo proposto de reconhecimento de faces invariante à iluminação.

Na tarefa de identificação de faces existem diferentes fatores que influenciam na performance do reconhecimento, entre os principais fatores encontra-se a variância de iluminação. No mundo real não é possível controlar a iluminação. Com isso e com a forma 3D da face, muitas imagens possuem visão parcial ocultando características relevantes para a identificação.

No desenvolvimento do modelo proposto foram realizadas a sequência de tarefas como apresentado no diagrama da Figura 4.1. Tendo o banco de imagens, este passa primeiramente por um pré-processamento onde é tratada a questão da variância de iluminação. Para esta tarefa existem duas possibilidades onde é realizada uma ou outra. A primeira é fazendo a normalização da variância de iluminação no domínio espacial utilizando o método LCE e a segunda é fazendo a normalização da variância de iluminação no domínio do DCT.

Após o pré-processamento para tratar a variação de iluminação, é feita a extração de características das imagens utilizando o DCT. Em seguida é reduzida a dimensionalidade empregando o método DPA. Esta tarefa pode ou não acontecer. E por fim é realizado o reconhecimento utilizando um classificador de aprendizado supervisionado.

As próximas seções descrevem os métodos utilizados no modelo proposto. As in-



**Figura 4.1:** Modelo Proposto de Reconhecimento de Faces Invariante a Iluminação.

formações foram retiradas dos trabalhos apresentados por (Kao, Hsu and Yang 2010, Dabbaghchian, Ghaemmaghani and Aghagolzadeh 2010, Chen, Meng and Shingian 2006).

## 4.2 Normalização da iluminação

Nesta seção é apresentado o pré-processamento do método proposto. Neste momento é tratada a questão da variação de iluminação para que a mesma se apresente estável e não interfira na classificação e conseqüentemente no reconhecimento da face. Para esta questão de variação de iluminação são utilizados por exemplo, equalização do histograma (HE), correção de gama, transformação logarítmico entre outros (Shan, Gao, Cao and Zhao 2003a, Savvides and Kumar 2003). No entanto ainda é complexo lidar com o uso dessas técnicas de processamento global em variação de iluminação não uniforme.

No modelo proposto para tratar a variação de iluminação, é feita uma normalização da mesma, de forma que possam ser retiradas características também das partes escuras da imagem de face. Esta normalização é realizada no presente trabalho de duas formas como é apresentado a seguir. A primeira é realizada no domínio espacial com o uso do LCE e a segunda é feita no domínio DCT.

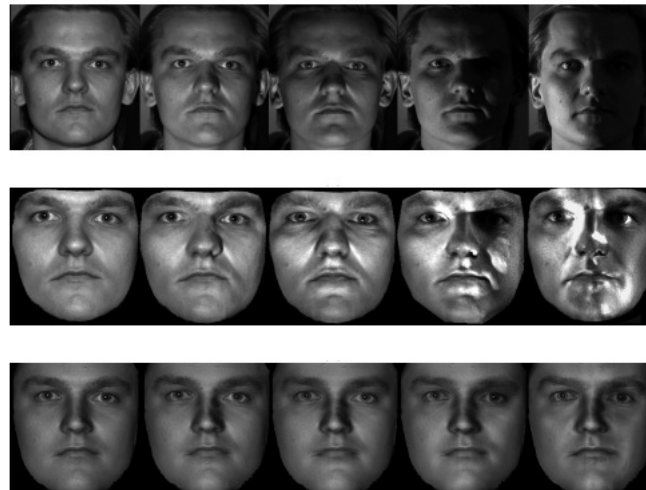
### 4.2.1 No domínio espacial (Realce do Contraste Local - LCE)

Dentro do reconhecimento de face, fazer a identificação de uma face em uma imagem com maior contraste, ou seja, a imagem deve possuir uma faixa dinâmica estendida, é uma questão crítica para a questão de desenvolvimento de câmeras de vigilância. A

câmera tradicional realiza exposição automática para determinar a exposição correta que cobre a faixa de intensidade da luz na imagem. Porém a faixa dinâmica de um sensor de imagem na câmera é muito menor do que a cena real. Sem compensar as variações de iluminação é impossível para um sistema de reconhecimento facial, alcançar taxas de reconhecimento aceitáveis (Lee, Ho and Kriegman 2005, Ishiyama, Hamanaka and Sakamoto 2005, Venkataramani, Qidwai and Vijayakumar 2005, Chen, Meng and Shingian 2006).

Existem duas formas de lidar com a variação de iluminação em reconhecimento facial que são a reiluminação do rosto e o realce da imagem como mostrado na Figura 4.2. Na figura, a primeira linha de imagens de face são as imagens originais com variação de iluminação. As linhas subsequentes representam os métodos propostos por (Wen, Liu and Huang 2003, Wang, Liu, Hua, Wen, Zhang and Samaras 2007).

Um algoritmo que trata a variação de iluminação com a reiluminação da face, sintetiza uma imagem sob qualquer condição de iluminação (Wang, Liu, Hua, Wen, Zhang and Samaras 2007, Shim, Luo and Chen 2008). Porém, geralmente, possui problema de complexidade maior, já que o problema é formulado como uma minimização de energia com algumas restrições.



**Figura 4.2:** Reiluminação de faces (Wang, Liu, Hua, Wen, Zhang and Samaras 2007).

Com isso, grande parte dos sistemas de reconhecimento invariante à iluminação adotam ferramentas de melhoria da imagem para preservação do contraste visual original e minimização dos artefatos indesejados, desta forma, fazendo o realce da imagem. Entre

as técnicas desenvolvidas para aumento de contraste está o LCE (Caselles, Lisani and Sapiro 1999, Kim and Hwang 2001, Gross and Brajovic 2003).

O LCE é eficiente para realce de visibilidade das finas texturas. Embora possa resultar em problemas de inversão de gradientes visíveis ou gerar halos indesejados, essas desvantagens podem ser eliminadas. A técnica é particularmente útil para melhorar o contraste dos detalhes da imagem HDR, já que os parâmetros extraídos são instáveis em imagens com fonte de luz desigual. Assim o LCE, primeiro calcula, para um dado  $pixel(x, y)$  com valor de iluminação  $Im(x, y)$  o contraste local  $\delta(x, y)$ , pela fórmula da Equação 4.1,

$$\delta(x, y) = \begin{cases} \log(Im(x, y)/\overline{Im(m, n)}) & \text{Se } Im(x, y) > \theta \text{ e } \overline{Im(x, y)} > \theta \\ 0 & \text{caso contrário} \end{cases} \quad (4.1)$$

onde  $\theta$  é pré-definido como 1.0 e  $\overline{Im(x, y)}$  denota a luminosidade média dos  $pixels$  vizinhos em uma janela de  $5 \times 5$  como indicada na Equação 4.2.

$$\overline{Im(x, y)} = \frac{1}{25} \sum_{i=-2}^2 \sum_{j=-2}^2 Y(x + i, y + j) \quad (4.2)$$

Ao invés de utilizar o valor da intensidade do  $pixel$ , o mesmo é representado pelo contraste local. Como os valores dos contrastes locais obtidos através da Equação 4.2 podem ser positivos ou negativos, então torna-se necessário a normalização dos dados. Assumindo que  $\delta_{max}$  e  $\delta_{min}$  representam o valor máximo e mínimo do contraste local de todos os  $pixels$ , respectivamente.

O valor de contraste local para um  $pixel(x, y)$  é fortemente normalizado pela Equação 4.3,

$$C(x, y) = \Phi \cdot \frac{(\delta(m, n) - \delta_{min})}{(\delta_{max} - \delta_{min})} \quad (4.3)$$

onde  $C(x, y)$  representa o valor do contraste local normalizado do  $pixel(x, y)$ , e o  $\Phi$  denota o máximo nível de cinza no intervalo de dados da imagem. Na Figura 4.3 é mostrado

o antes e o depois da aplicação da técnica LCE.



**Figura 4.3:** Antes e depois da aplicação do LCE (Kao, Hsu and Yang 2010)

#### 4.2.2 Normalização da iluminação no domínio DCT - Transformada Logarítmica

A transformada logarítmica, frequentemente, é utilizada para realçar imagens expandindo os valores de *pixels* escuros (Adini, Moses and Ullman n.d., Gonzalez and Woods 1992). A compensação da iluminação pode ser implementada no domínio logarítmico. Uma imagem em níveis de cinza  $Im(x, y)$  pode ser proporcional ao produto da reflectância  $r(x, y)$  com a iluminação  $e(m, n)$  (Horn 1986) como na Equação 4.4.

$$Im(x, y) = r(x, y) * e(x, y) \quad (4.4)$$

Como a reflectância é uma característica estável da face, o objetivo é recuperar

a mesma sob condições de iluminação variadas. Para isso é tomada a transformada logarítmica (4.4) e obtém-se a Equação 4.5.

$$\log Im(x, y) = \log r(x, y) + \log e(x, y). \quad (4.5)$$

A partir da Equação 4.5 podemos determinar que no domínio logarítmico, se a iluminação incidente é  $e(m, n)$  e a iluminação uniforme desejada  $e'$  são dadas ( $e'$  é idêntica para cada *pixel* da imagem) obtém-se a Equação 4.6,

$$\begin{aligned} \log Im'(x, y) &= \log r(x, y) + \log e' \\ &= \log r(x, y) + \log e(x, y) - \epsilon(x, y) \\ &= \log Im(x, y) - \epsilon(x, y) \end{aligned} \quad (4.6)$$

onde

$$\epsilon(x, y) = \log e(x, y) - \log e' \quad (4.7)$$

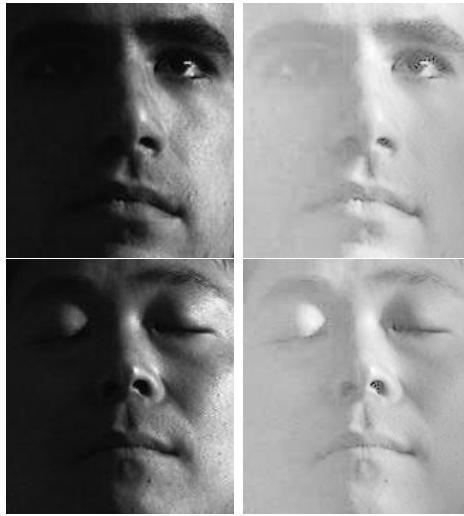
e  $Im'(x, y)$  é o valor do *pixel* com a iluminação uniforme desejada. Devido a Equação 4.6 pode-se concluir que a normalização da imagem pode ser obtida da imagem original usando o termo  $\epsilon(m, n)$  que é chamado de termo de compensação que é a diferença entre a iluminação normalizada e a iluminação original estimada no domínio logaritmo.

### Compensação da iluminação

Em uma imagem com variações de iluminação, esta pode ser compensada somando e subtraindo o termo de compensação  $\epsilon(x, y)$  de (Equação 4.6) no domínio logarítmico. A variação de iluminação em imagens de face muda lentamente em comparação a reflectância. Uma vez que as variações de iluminação residem, principalmente, na banda de baixa frequência, estas podem ser reduzidas através da remoção dos componentes de baixa frequência.

O DCT pode ser utilizado para transformar uma imagem a partir do domínio espacial para o domínio da frequência. Os componentes de baixa frequência podem ser removidos fixando-os em zero. Assim o sistema resultante funciona como um filtro passa-alta.





**Figura 4.4:** Exemplo de compensação da iluminação no domínio DCT.

Isso pode ser observado na Equação 4.6, onde a fixação dos coeficientes do DCT para zero é equivalente a subtrair o produto da imagem com base no DCT e do coeficiente correspondente a partir da imagem original. Se  $n$  for um coeficiente de baixa frequência então é atribuído o valor zero e assim tem-se 4.8,

$$\begin{aligned}
 Im'(x, y) &= \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} E(u, v) - \sum_{i=1}^n E(u_i, v_i) \\
 &= MC(x, y) - \sum_{i=1}^n E(u_i, v_i)
 \end{aligned} \tag{4.8}$$

onde  $MC(x, y)$  é a transformada DCT e

$$E(u, v) = \frac{1}{\sqrt{PQ}} \alpha(u) \alpha(v) \sum_{x=0}^{P-1} \sum_{y=0}^{Q-1} Im(x, y) \times \cos\left(\frac{(2x+1)u\pi}{2P}\right) \times \cos\left(\frac{(2y+1)v\pi}{2Q}\right) \tag{4.9}$$

O termo  $\sum_{i=1}^n E(u_i, v_i)$  é o termo de compensação da iluminação e  $Im'(x, y)$  é a face normalizada no domínio logaritmo. Na Figura 4.4 pode ser observado como fica a imagem resultante após realizar a compensação da iluminação.

O primeiro coeficiente DCT, isto é, o componente DC, determina a iluminação da imagem da face. Assim pode-se alcançar uma iluminação uniforme ajustando o mesmo. Isso pode ser obtido através da Equação 4.10,

$$C(0, 0) = \log \mu \sqrt{PQ} \quad (4.10)$$

onde  $C(0, 0)$  é o coeficiente logarítmico DC de imagem.

Para uma melhor compreensão e visualização Chen, Meng and Shingian (2006) sugerem que o valor de  $\mu$  seja próximo do valor médio das intensidade da imagem original, em outras palavras, a face normal tem um nível médio de cinza. Uma observação é que para questões de iluminação não é levado em consideração a cor da pele como característica, pois a mesma é instável. Assim, resulta a partir das Equações 4.8 e 4.6 que a diferença entre o componente DC original e do componente DC normalizado, juntamente com os outros componentes de baixa frequência AC descartados, se aproxima do termo de compensação  $\epsilon(m, n)$ .

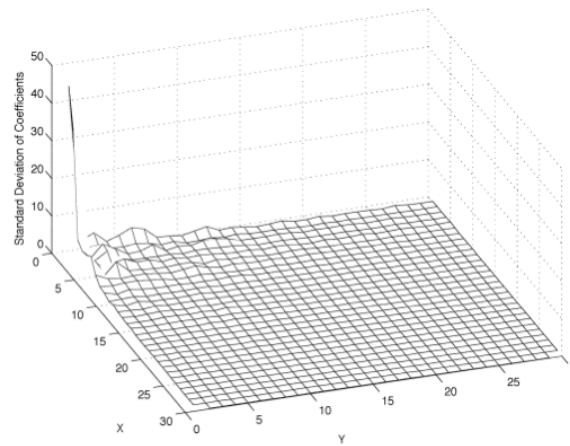
### Descartando coeficientes DCT

Os coeficientes DCT de baixa frequência são altamente ligados a variações de iluminação e devem ser descartados. Porém, uma questão é estimar quais e quantos coeficientes devem ser descartados a fim de obter a melhor imagem da face normalizada.

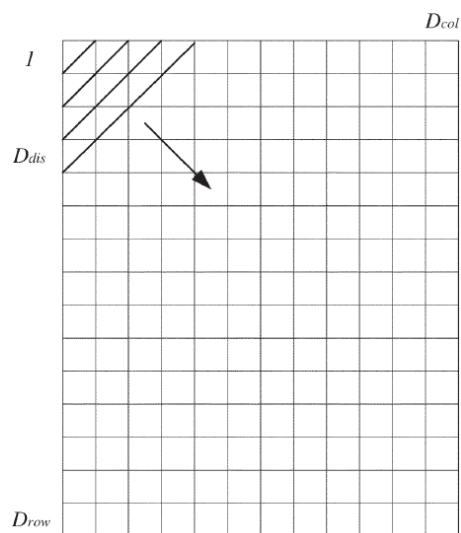
Na Figura 4.5 é mostrado os desvios-padrão dos coeficientes DCT calculados a partir de imagens de face do mesmo tipo. Como pode ser observado os coeficientes de grande magnitude estão localizados principalmente no canto superior esquerdo da matriz de coeficientes DCT. Com isso, a variação de iluminação pode ser amenizada descartando esses coeficientes de baixa frequência e a forma utilizada para descartar os coeficientes DCT é apresentada na Figura 4.6.

No trabalho apresentado por Chen, Meng and Shingian (2006) é proposta uma nova abordagem de normalização da informação. O mesmo reduz significativamente a variação de iluminação descartando os coeficientes DCT de baixa frequência no domínio logarítmico. Além disso, o modelo possui entre outras vantagens rapidez, facilidade de implementação e uma melhor performance do que as abordagens existentes. O autor sugere que o número de coeficientes DCT descartados seja no intervalo de 18 a 25 coeficientes. Neste trabalho foram descartados o número intermediário de 20 coeficientes de baixa frequência.

Após serem retirados os coeficientes DCT da imagem, é então calculado a transfor-



**Figura 4.5:** Desvio padrão dos coeficientes DCT (Chen, Meng and Shingian 2006).



**Figura 4.6:** Descartando coeficientes DCT (Chen, Meng and Shingian 2006).

mada inversa DCT obtendo assim a imagem normalizada. Posteriormente essa imagem será utilizada na fase de extração de características.

### 4.3 Extração de características

Em Chen, Meng and Shingian (2006) após descartar os  $n$  primeiros coeficientes DCT é então realizado o inverso DCT, em seguida, o processo de reconhecimento é executado diretamente na imagem de domínio logaritmo, e a transformada logarítmica inversa é ignorada. Os autores mostraram que o PCA pode ser feito no domínio DCT e os mesmos resultados são obtidos como se fosse aplicado no domínio espacial. Neste trabalho, a extração de características é executada diretamente no domínio DCT.

Existem duas maneiras distintas de extrair as características utilizando o DCT. Na primeira, a imagem é utilizada como um único bloco onde o DCT é aplicado para obter os coeficientes. Na segunda maneira, a imagem é dividida em pequenos blocos e o DCT é aplicado em cada um dos blocos.

Aqui é usada uma abordagem baseada em blocos, onde a imagem é dividida em blocos regulares de  $8 \times 8$ . Portanto, cada bloco contém 64 coeficientes DCT.

Alguns dos componentes de baixa frequência são geralmente selecionados como valores de características para classificação. Como sugerido por Kao, Hsu and Yang (2010), são utilizados os coeficientes DC (do inglês *discriminant coefficient*) que representam os coeficientes de maior magnitude. Estes, por sua vez, são organizados em um vetor e então usados para descrever o conteúdo da imagem. As características são selecionadas de todos os coeficientes DCT dos blocos particionados.

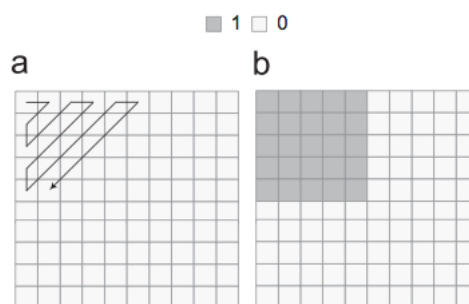
Apenas alguns componentes de baixa frequência são geralmente selecionados como características para a classificação. A seleção dos coeficientes DCT é uma tarefa importante do processo de extração de características. A maioria das abordagens baseadas no DCT, não dão atenção suficiente à seleção de coeficientes. Os coeficientes são geralmente escolhidos com métodos convencionais (mascaramento; ziguezague; ou zonal). Porém estas abordagens não são necessariamente eficazes em todas as aplicações.

## 4.4 Redução de dimensionalidade

Após a obtenção dos coeficientes extraídos no passo anterior, é realizada a seleção dos coeficientes de melhor representação da imagem e conseqüentemente reduzir a dimensionalidade do descritor. O objetivo de aplicar a redução de dimensionalidade é analisar se a uma melhora da eficácia quando é utilizado somente os coeficientes que melhor representam a imagem.

Selecionar os coeficientes de melhor representação da imagem é uma parte importante no processo de extração de características e influencia fortemente a precisão do reconhecimento. Pan, Rust and Bolouri (2000) propuseram um método para selecionar os coeficientes que minimizam o erro de construção. Porém este é um bom critério para compressão e não para reconhecimento. Jing and Zhang (2004) apresentaram uma nova abordagem para encontrar bandas discriminantes no espaço transformado. Sanderson and Paliwal (2003) utilizam os coeficientes do polinômio derivado a partir dos coeficientes 2D-DCT obtidos em blocos espacialmente vizinhos.

As abordagens citadas selecionam elementos fixos na matriz de coeficientes DCT que são chamadas de abordagens determinísticas por (Dabbaghchian, Ghaemmaghami and Aghagolzadeh 2010). Ziguezague, máscara zonal e suas modificações são exemplos das abordagens determinísticas que podem ser observadas na Figura 4.7.



**Figura 4.7:** (a) Abordagem determinística Ziguezague; (b) Abordagem determinística Máscara Zonal (Dabbaghchian, Ghaemmaghami and Aghagolzadeh 2010)

Apesar das abordagens determinísticas serem simples, não são necessariamente eficientes em todas às bases de dados. Utilizar o DCT em um vetor de características, o que melhora o desempenho de reconhecimento em uma banco de dados, pode deteriorar os resultados em outra. Em outras palavras, o DCT não apresenta bons resultados para todas às situações.

Assim é proposto um novo grupo de abordagem estatística chamado Dados Dependentes (do inglês *Data-dependent*) por (Dabbaghchian, Ghaemmaghami and Aghagolzadeh 2010). A ideia principal se baseia no fato de que todos os coeficientes DCT não possuem a mesma capacidade de discriminar diferentes classes. Em outras palavras, alguns coeficientes separam melhor as classes do que outros. Assim o objetivo é encontrar estes coeficientes no banco de dados.

Para isso a proposta é um método de análise estatística, onde são exploradas todas as imagens do banco de dados para associar cada coeficiente DCT a um número que representa o seu poder de discriminação (DP). Esta técnica é chamada de DPA. A ideia de dados dependentes e DPA podem ser implementados por qualquer problema de seleção de recursos, bem como os coeficientes DCT.

O DPA é diferente de outros métodos semelhantes, tais como PCA e LDA, que também utilizam variações intra-classe e inter-classes. Enquanto que abordagens tais como PCA e LDA tentam obter uma transformação que maximiza a discriminação das características no domínio transformado, o DPA procura pelas melhores características no domínio original. Além disso o DPA não possui nenhum problema de singularidade e pode ser usado como algoritmos de redução de características e combinado com outros algoritmos.

Um coeficiente discriminante depende de dois atributos, o primeiro é a grande variação entre as classes, ou seja, características que diferenciam duas classes e o segundo é a pequena variação dentro das classes, ou seja, características que definem uma classe. Quanto mais definidos esses dois atributos maior o poder de discriminação. Assim o coeficiente discriminante pode ser obtido pela divisão da variância entre-classe pela variância intra-classe. Com isso, espera-se gerar valores altos para os coeficientes discriminantes.

Considerando que a matriz de coeficientes DCT de uma imagem  $Im(x, y)$  é representada por MC mostrada na Equação 4.11.

$$MC = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1y} \\ c_{21} & c_{22} & \dots & c_{2y} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ c_{x1} & c_{x2} & \dots & c_{xy} \end{bmatrix}_{x \times y} \quad (4.11)$$

Para calcular o DP dos coeficientes  $c_{ij}$  ( $i = 1, 2, \dots, x$  e  $j = 1, 2, \dots, y$ ), tendo um número  $T$  de classes e um número  $S$  de imagens de treinamento de cada classe, é construída a matriz de treino  $A_{ij}$  mostrada na Equação 4.12 onde é escolhido o coeficiente DCT das posições  $i$  e  $j$  para todas as classes e para todas imagens de treino.

$$A_{ij} = \begin{bmatrix} c_{ij}(1, 1) & c_{ij}(1, 2) & \dots & c_{ij}(1, T) \\ c_{ij}(2, 1) & c_{ij}(2, 2) & \dots & c_{ij}(2, T) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ c_{ij}(S, 1) & c_{ij}(S, 2) & \dots & c_{ij}(S, T) \end{bmatrix}_{S \times T} \quad (4.12)$$

Em seguida é calculado o valor médio de cada classe com a equação 4.13.

$$M_{ij}^t = \frac{1}{S} \sum_{s=1}^S A_{ij}(S, T), \quad t = 1, 2, \dots, T \quad (4.13)$$

Depois é gerada a variância de cada classe através da Equação 4.14.

$$V_{ij}^t = \sum_{s=1}^S (A_{ij}(S, T) - M_{ij}^t)^2, \quad (4.14)$$

$$t = 1, 2, \dots, T \quad (4.15)$$

$$(4.16)$$

Então é obtida a variação média de todas as classes através da Equação 4.17.

$$V_{ij}^W = \frac{1}{T} \sum_{t=1}^T V_{ij}^t \quad (4.17)$$

Em seguida é calculado a média de todas as amostras de treinamento pela Equação 4.18.

$$M_{ij} = \frac{1}{S \times T} \sum_{t=1}^T \sum_{s=1}^S A_{ij}(S, T) \quad (4.18)$$

Depois é computado também, a variância de todas as amostras de treinamento pela equação 4.19.

$$V_{ij}^B = \sum_{t=1}^T \sum_{s=1}^S (A_{ij}(S, T) - M_{ij})^2 \quad (4.19)$$

E por fim é estimado o DP pela localização  $(i, j)$  com a Equação 4.20

$$D_{ij} = \frac{V_{ij}^B}{V_{ij}^W}, \quad (4.20)$$

$$1 \leq i \leq x, \quad (4.21)$$

$$1 \leq j \leq y \quad (4.22)$$



Com isso, quanto maior for o valor de DP maior é a capacidade de discriminação do coeficiente correspondente. Em outras palavras, espera-se alcançar a taxa de reconhecimento máxima utilizando os coeficientes que possuem poder de discriminação máximo.

No DPA, em se tratando de uma abordagem estatística, a quantidade de imagens de treinamento afeta os valores DPs. Este fato acontece com mais intensidade para os coeficientes de baixa frequência do que para os outros.

## 4.5 Classificação

Neste trabalho o conhecimento desejado é a classificação das características específicas da imagem de uma face de forma que identifiquem a que pessoa pertence. Para obter esse conhecimento pode-se utilizar técnicas existentes de aprendizado de máquina, como a classificação com aprendizado supervisionado.

Para a realização do reconhecimento das faces, foi escolhido o classificador SVM linear (LIBSVM) utilizando o kernel Gaussiano (Chang and Lin 2001). Esta é uma biblioteca muito utilizada entre os trabalhos apresentados na literatura. Este realiza o reconhecimento de padrões separando as características da face em classes como foi discutido no Capítulo 3.

## 4.6 Considerações finais

O modelo proposto apresentado neste capítulo tem como propriedade o reconhecimento de face com o tratamento da variação de iluminação. Para isso é feito um pré-processamento nas imagens de face que pode ser através de dois métodos. O primeiro é a normalização da iluminação no domínio espacial com a técnica LCE que é uma técnica eficiente para o melhoramento da visibilidade das finas texturas e muito útil para melhorar o contraste dos detalhes. O segundo método é a normalização da iluminação no domínio DCT através da transformada logarítmica.

Após a utilização de uma das técnicas para o pré-processamento das imagens, é feita então a extração de características também através da técnica DCT que é aplicada, agora, na imagem normalizada. Em seguida pode ser empregado ou não a redução de

dimensionalidade nos vetores de características selecionando os coeficientes de melhor representação da imagem. Para isso é utilizada a técnica DPA onde é atribuído para cada coeficiente um número que representa o poder discriminativo. E por fim é feito o reconhecimento através do classificador SVM.

No próximo capítulo são apresentados os bancos de dados utilizados para os testes e como foram feitos os experimentos. É também expostos os resultados obtidos para as variações do modelo proposto que são: com o uso do LCE ou DCT para o pré-processamento e com ou sem o uso do DPA para a redução de dimensionalidade. Além disso são mostrados os resultados dos métodos apresentados na literatura.

# Capítulo 5

## Resultados

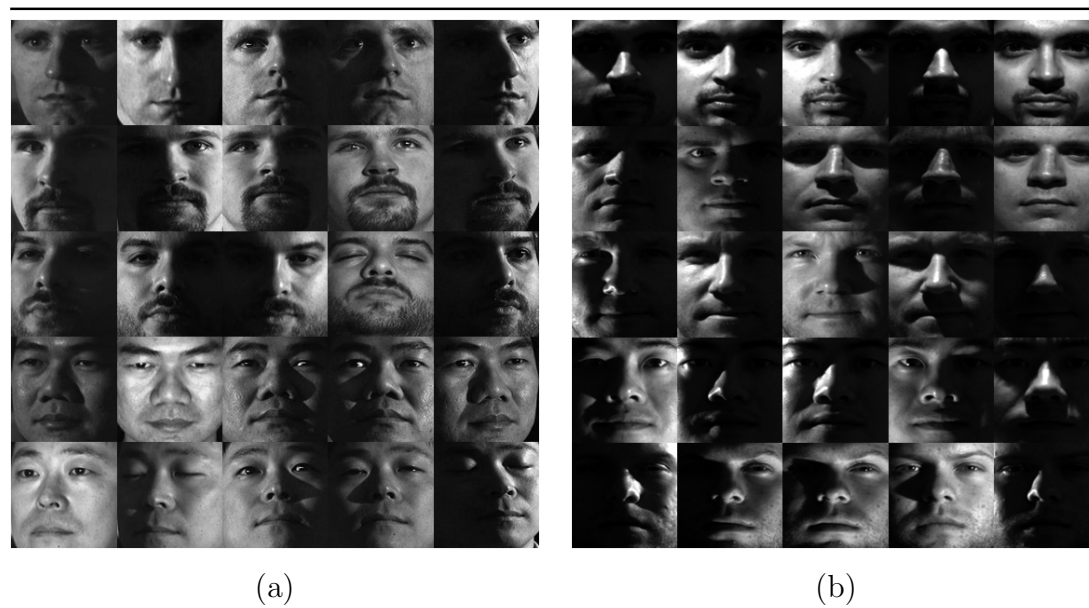
Para uma melhor análise do desenvolvimento do modelo proposto do presente trabalho, foram realizados três experimentos. No primeiro, foi feita a comparação das variações do modelo proposto, ou seja, é analisado o desempenho dos métodos utilizando o LCE ou o DCT para o pré-processamento da imagem e também com ou sem o DPA para a redução de dimensionalidade para verificar qual deles é mais eficaz. Para esta avaliação são utilizados os bancos de dados CMU PIE e YaleB (Georghiades, Belhumeur and Kriegman 2001a) ambos com imagens de face com grandes variações de iluminação.

No segundo experimento, foi comparado o método proposto com outros métodos utilizados quando há variação de iluminação. Para este experimento foram utilizados também os dois bancos de dados: Yale B e o CMU PIE.

E no terceiro experimento é avaliado o desempenho da abordagem proposta com diferentes configurações, vista frontal e de perfil e imagens faciais com fundo confuso (não uniforme). Para este experimento foram utilizados quatro bancos de dados variados que foram *Japanese Female Facial Expression* (JAFFE) (Kamachi, Lyons and Gyoba 1998); AT&T (Cambridge 2002); UMIST (Graham and Allinson 1998) e Georgia (*Georgia Tech Face Database* 2007).

### 5.1 Descrição dos bancos de dados

O banco de dados CMU PIE é composto por 68 pessoas, com 13 posições distintas, 43 condições de iluminação diferentes e com 4 expressões desiguais. A YaleB é composta por



**Figura 5.1:** Exemplo de imagens das bases de dados (a) PIE e (b) Yale.

2432 imagens de 38 pessoas, com diferentes expressões faciais e condições de iluminação. A Figura 5.1 apresenta exemplos das bases de dados PIE e YaleB.

O banco de dados JAFFE contém um total de 213 imagens com 16 pessoas com 7 imagens de diferentes expressões de cada pessoa, sendo 6 de expressões faciais e uma neutra. AT&T também conhecida como a banco de dados ORL é composta de 400 imagens de 40 indivíduos, algumas imagens são capturadas em diferentes intervalos de tempo e tem variações tais como expressões e detalhes faciais, como também há inclinação e rotação de face de até  $20^\circ$ . Na Figura 5.2 são apresentados exemplos de imagens das bases de dados JAFFE e AT&T. O banco de dados *Sheffield* (antigamente conhecida como UMIST) consiste em 564 imagens de 20 indivíduos nos quais apresentam diferentes poses de frente e de perfil, o que dificulta a tarefa de reconhecimento. E por fim o banco de dados Georgia que contém imagem de 50 pessoas representadas por 15 imagens coloridas com fundo confuso. Exemplos de imagens das bases Georgia e UMIST são apresentados na Figura 5.3.

Como foi descrito anteriormente todos os bancos de dados consistem em um conjunto de imagens de pessoas diferentes. Cada grupo de imagens de uma determinada pessoa é uma classe na qual o sistema, através das características extraídas, classifica a imagem. A tabela 5.1 apresenta um comparativo das características dos bancos de dados utilizados nos testes para uma melhor visualização dos diferentes pontos testados.

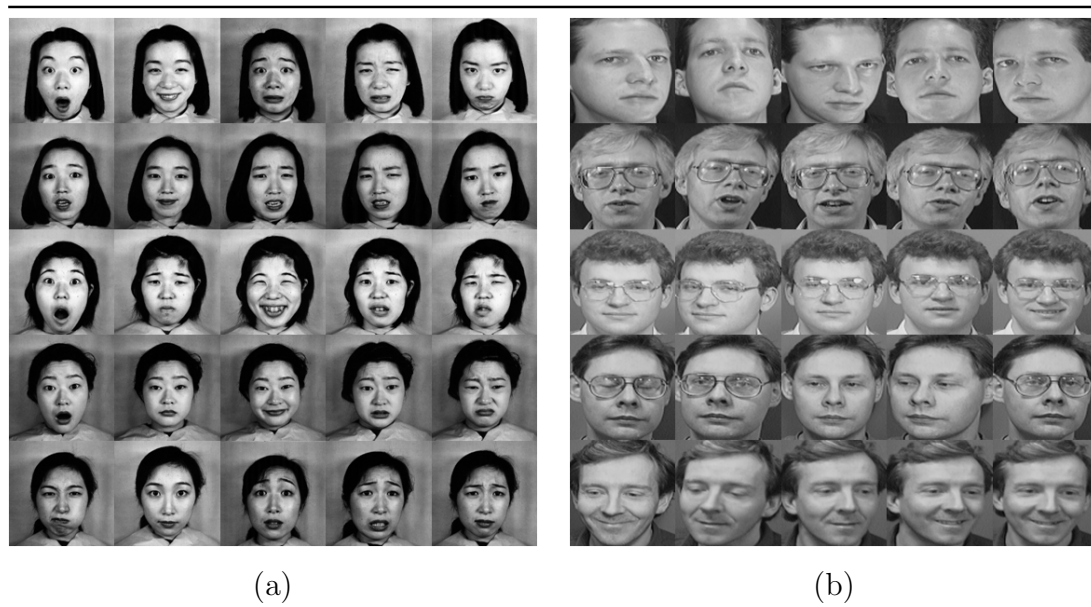


Figura 5.2: Exemplo de imagens das bases de dados (a) JAFFE e (b) AT&T.

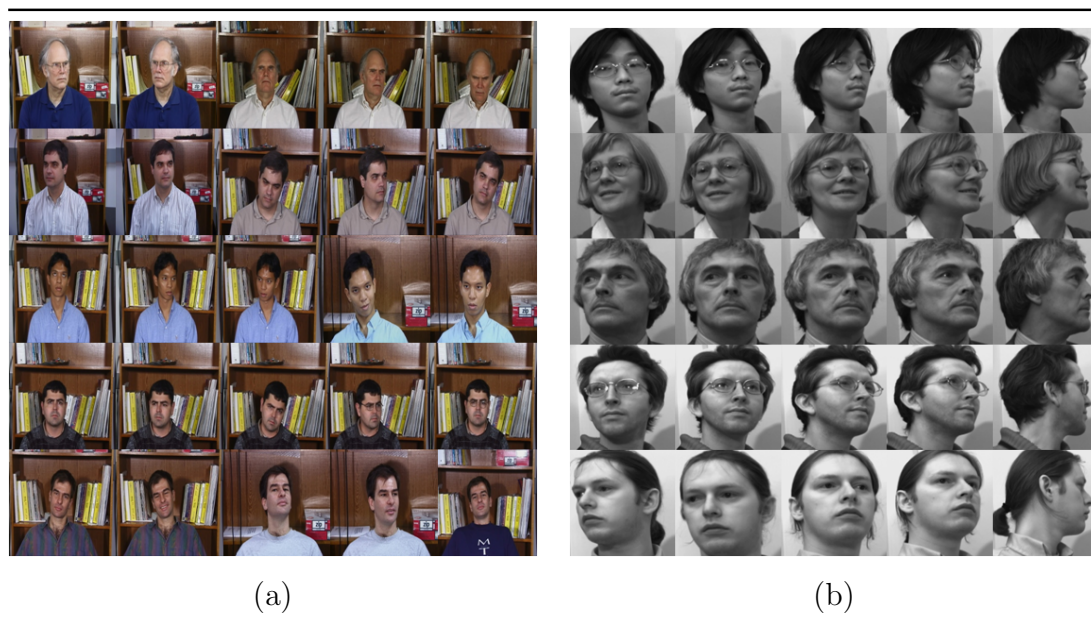


Figura 5.3: Exemplo de imagens das bases de dados (a) Georgia e (b) UMIST.

**Tabela 5.1:** Características dos bancos de dados utilizados

Banco de Dados	Imagens	Classes	Características
CMU PIE	884	68	condições de iluminação distintas; variação de expressão
YaleB	2432	38	condições de iluminação distintas; variação de expressão
JAFFE	213	16	variação de expressão
AT&T	400	40	variação de expressão; feitas em tempos distintos; inclinação; rotação
UMIST	564	20	variação de poses (frente e perfil)
Georgia	750	15	coloridas; variação de fundo

## 5.2 Protocolo de experimentos

Para uma melhor confiabilidade dos resultados, cada método foi executado utilizando validação cruzada com cinco *folds*, ou seja, os testes foram realizados cinco vezes sendo tirado a média dos resultados. O protocolo seguido para avaliação dos experimentos foram conduzidos em quatro medidas: acurácia, precisão, revocação e variância. A acurácia representa a porcentagem de acertos do modelo proposto em cada banco. A precisão representa a fração de itens reais positivos que foram classificados corretamente. Já a revocação representa a quantidade que foi classificada com sucesso, isto é, quantos itens foram corretamente classificados como uma classe positiva.

As equações utilizadas para calcular a acurácia, precisão e revocação são as Equações 5.1, 5.2 e 5.3 respectivamente.

$$Acc = \frac{VerdadeiroPositivo + VerdadeiroNegativo}{VerdadeiroPositivo + FalsoNegativo + VerdadeiroNegativo + FalsoPositivo} \quad (5.1)$$

$$P = \frac{VerdadeiroPositivo}{VerdadeiroPositivo + FalsoPositivo} \quad (5.2)$$

$$R = \frac{\textit{VerdadeiroPositivo}}{\textit{VerdadeiroPositivo} + \textit{FalsoNegativo}} \quad (5.3)$$

onde Verdadeiro Positivo é a proporção de verdadeiros positivos em relação a todas as predições positivas e Verdadeiro Negativo é a proporção de verdadeiros negativos em relação a todas predições negativas.

A variância é a medida de dispersão ou variabilidade em uma distribuição (Montgomery and Runger 2003), ou seja, aqui neste trabalho, é a medida da dissipação dos cinco testes feitos para cada banco de dados. A variância é denotada por  $\sigma^2$  ou  $V(X)$  e é calculada a partir da Equação 5.4,

$$\sigma^2 = V(X) = E(X - \mu)^2 = \sum_x (x - \mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2 \quad (5.4)$$

onde  $\mu$  é a média dos resultados.

### 5.3 Experimento I

O modelo proposto desenvolvido no presente trabalho objetiva fazer o reconhecimento de faces em imagens com variação de iluminação. Dois trabalhos, apresentados na literatura, também são utilizados para o reconhecimento de face em imagens escurecidas. O primeiro é o método apresentado por Chen, Meng and Shingian (2006). O mesmo reduz a variação de iluminação truncando os coeficientes de baixa frequência da transformada DCT no domínio logarítmico. O segundo trabalho foi desenvolvido por Kao, Hsu and Yang (2010). Este, por sua vez, utiliza o aumento do contraste local para reduzir as variações de iluminação.

No Experimento I é testado qual as variações do modelo proposto apresenta um melhor desempenho. Existem quatro variações do modelo: a primeira é utilizando o LCE para pré-processamento da imagem e o DCT para extração de característica (LCE+DCT); a segunda, utiliza o LCE para pré-processamento da imagem, o DCT para extração de característica e o DPA para redução de dimensionalidade (LCE+DCT+DPA); a terceira utiliza a normalização no domínio DCT para pré-processamento da imagem e também o DCT para extração de característica (DCT+DCT) e a quarta, a normalização no domínio DCT para pré-processamento da imagem, o DCT para extração de

característica e o DPA para redução de dimensionalidade (DCT+DCT+DPA). Para uma melhor observação dos métodos que contém em cada variação do modelo proposto pode-se observar a tabela 5.2.

**Tabela 5.2:** Variações do modelo proposto no presente trabalho

Modelos	Normalização (LCE)	Normalização (DCT)	Extração de características (DCT)	Redução de dimensionalidade (DPA)
VariaçãoI (LCE+DCT)	sim	—	sim	—
VariaçãoII (LCE+DCT+DPA)	sim	—	sim	sim
VariaçãoIII (DCT+DCT)	—	sim	sim	—
VariaçãoVI (DCT+DCT+DPA)	—	sim	sim	sim

Na Tabela 5.3 é apresentado a acurácia e a variância para comparação das versões do modelo proposto. Para esta análise comparativa foram utilizados os bancos de dados CMU PIE e Yale B, ambos com imagens com grandes variações de iluminação.

**Tabela 5.3:** Resultados das variações do modelo proposto LCE+DCT, LCE+DCT+DPA, Normalização no domínio DCT+DCT e Normalização no domínio DCT+DCT+DPA. A variável (Acc) representa a acurácia e a variável (Var) a variância.

Banco de Dados	LCE+DCT		LCE+DCT+DPA		DCT+DCT		DCT+DCT+DPA	
	Acc	Var	Acc	Var	Acc	Var	Acc	Var
CMU PIE	98,701	0,198	98,028	0,101	<b>99,875</b>	<b>0,012</b>	<b>99,962</b>	<b>0,008</b>
YaleB	95,535	0,626	97,112	0,249	<b>99,875</b>	<b>0,026</b>	<b>99,938</b>	<b>0,017</b>

Na Tabela 5.4 é apresentado a comparação das versões do modelo proposto, porém levando em consideração a precisão e a revocação . Para esta análise comparativa também foram utilizados os bancos de dados CMU PIE e Yale B.



**Tabela 5.4:** Resultados das variações do modelo proposto LCE+DCT, LCE+DCT+DPA, Normalização no domínio DCT+DCT e Normalização no domínio DCT+DCT+DPA. A variável (Prec) representa a precisão e a variável (Rev) a revocação.

Banco de Dados	LCE+DCT		LCE+DCT+DPA		DCT+DCT		DCT+DCT+DPA	
	Prec	Rev	Prec	Rev	Prec	Rev	Prec	Rev
CMU PIE	0,987	0,987	0,980	0,980	<b>0,999</b>	<b>0,999</b>	<b>1,000</b>	<b>1,000</b>
YaleB	0,959	0,955	0,972	0,971	<b>0,999</b>	<b>0,999</b>	<b>0,999</b>	<b>0,999</b>

## 5.4 Experimento II

No Experimento II é analisado o modelo proposto baseado no DCT para o pré-processamento de imagens com outros métodos utilizados também para o reconhecimento de faces invariantes a iluminação. A escolha da versão do modelo proposto é devido ao melhor desempenho apresentado como pode ser observado na Seção 5.3.

Na Tabela 5.5 é apresentado a acurácia e a variância para comparação dos métodos da literatura com as duas formas do modelo proposto: (Modelo Proposto+DPA) com a redução da dimensionalidade e (Modelo Proposto) sem o DPA. Ainda na Tabela 5.5 pode ser observado que o método proposto com e sem redução de dimensionalidade obtém resultados superiores em comparação com os outros dois métodos. Na Tabela 5.6 é mostrado para a comparação dos métodos em termos da precisão e revocação. Novamente, os métodos apresentados no presente trabalho, com e sem redução de dimensionalidade, alcançam melhores resultados.

Na Figura 5.4 e 5.5, são apresentadas em forma de gráfico os resultados mostrados nas tabelas 5.5 e 5.6 respectivamente. Nos gráficos pode ser observado que os métodos apresentados com e sem a redução de dimensionalidade obtém a melhor performance.

## 5.5 Experimento III

A fim de avaliar a abordagem proposta em situações distintas com outras configurações tais como variação de posição, de fundo, expressões e gestos, os experimentos foram

**Tabela 5.5:** Resultado dos experimentos considerando a acuracia e a variância. A acuracia é representada por (Acc) e a variância por (Var) para as bases de dados YaleB e CMU PIE.

Método	Yale B		PIE	
	Acc	Var	Acc	Var
Modelo proposto	99,875	0,026	99,923	0,012
Modelo proposto com DPA	<b>99,938</b>	0,017	<b>99,962</b>	0,008
(Chen, Meng and Shingian 2006)	96,312	0,004	95,907	0,016
(Kao, Hsu and Yang 2010)	96,730	0,163	95,647	0,018

**Tabela 5.6:** Resultado dos experimentos em termos da precisão e revocação. A precisão é representada por (Prec) e a revocação por (Rev) para as bases de dados Yale B e CMU PIE.

Method	Yale B		PIE	
	Prec	Rev	Prec	Rev
Modelo proposto	0,999	0,999	0,999	0,999
Modelo proposto com DPA	<b>0,999</b>	<b>0,999</b>	<b>1,000</b>	<b>1,000</b>
(Chen, Meng and Shingian 2006)	0,966	0,963	0,963	0,959
(Kao, Hsu and Yang 2010)	0,961	0,963	0,993	0,993

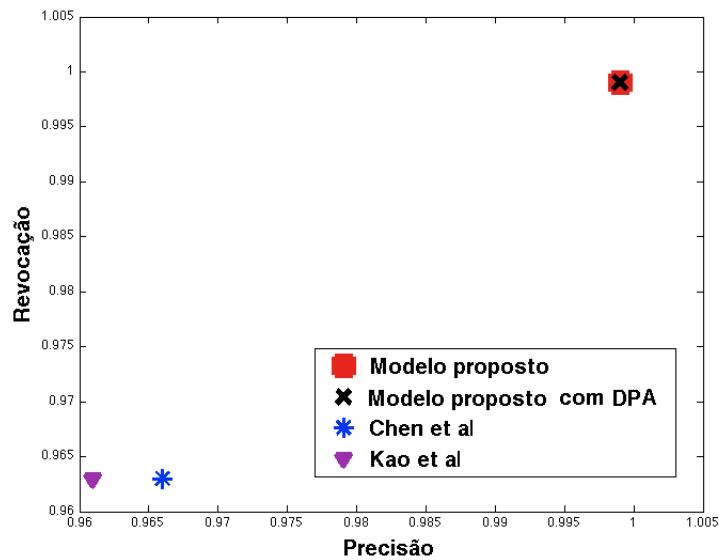


Figura 5.4: Gráfico comparativo para a banco de dados YaleB.

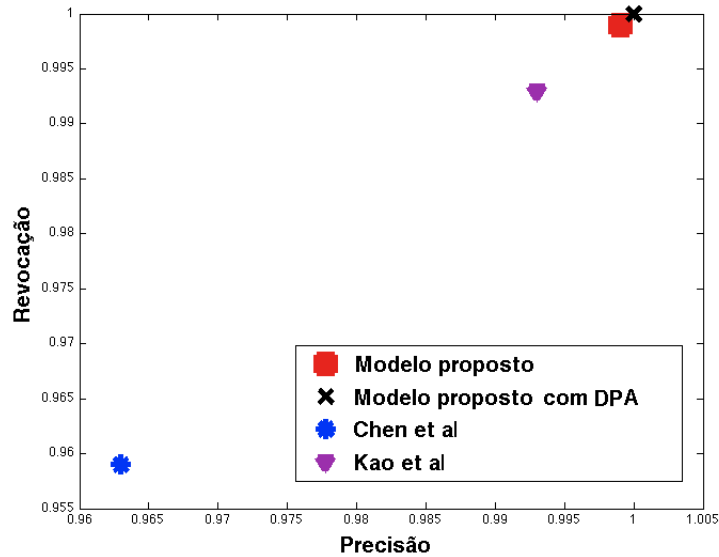


Figura 5.5: Gráfico comparativo para a banco de dados PIE.

realizados para testes em quatro conjuntos de dados de face utilizados na literatura. As bases utilizadas foram JAFFE, AT&T, UMIST e Georgia e estão descritos no início do capítulo. A abordagem proposta foi comparada com métodos apresentados na literatura.

Os primeiros métodos testados para comparação com o modelo proposto foram Ein-

geface e Fisherface. Os mesmos foram utilizados através da biblioteca *cvprtoolbox*<sup>1</sup>. O método SIFT foi utilizado a biblioteca *Vl-feat* (Vedaldi and Fulkerson 2008) e os métodos Mel-cepstrum e Mellin-cepstrum, foram implementados a partir dos artigos (Cakir and Cetin 2011).

Na tabela 5.7 é apresentado os resultados dos métodos implementados com o método proposto para comparação. Nesta tabela é levado em consideração a precisão e a revocação para as bases de dados JAFFE, AT&T, UMIST e Georgia. A variável  $P$  representa a precisão de cada resultado e a variável  $R$  a revocação. Mais uma vez, como pode ser observado, o desempenho obtido pelo método proposto no presente trabalho é melhor do que os métodos da literatura: Fisherface, SIFT, Mel e Mellin para todas as bases de dados. O método Eigenface obteve um melhor resultado com o banco de dados UMIST, porém com um pequena diferença. Vale ressaltar que o Eigenface é baseado na abordagem holística que possui como propriedade utilizar toda a imagem para o processo de reconhecimento deixando o método com um maior custo computacional.

O desempenho com mais variações ocorre com o banco de dados Georgia. Isto acontece devido ao fato de que este conjunto de dados não só tem variações na expressão de iluminação e posição, como também possui variações no fundo das imagens de face. Apesar destas variações, pode ser observado que a abordagem proposta obtém bons resultados em comparação com outros métodos.

O pior resultado para as bases de dados JAFFE e Georgia foi obtido por Fisherface que apresenta uma maior instabilidade entre os resultados. Para o banco de dados AT&T o pior resultado foi obtido por Melin-cepstrum e para a base UMIST o pior método foi o Mel. No caso da banco de dados AT&T, esta possui propriedades mais complexas por ser composta de imagens com diferentes expressões e também em espaços de tempos diferentes. Esta última propriedade permite uma variação muito grande das imagens com características tais como roupas, cabelo, barba entre outras.

Para uma melhor análise comparativa do método proposto, na tabela 5.8, é apresentado os resultados dos métodos da literatura para os bancos de dados JAFEE, AT&T, UMIST e Georgia, porém, aqui, é levado em consideração a acurácia e a variância obtidas por cada abordagem.

A abordagem proposta novamente atingiu os melhores resultados em três de quatro bancos de dados usados nos testes. A exatidão foi maior para os bancos de dados JAFFE,

---

<sup>1</sup><http://note.sonots.com/Matlab/cvprtoolbox.html>

**Tabela 5.7:** Resultados dos métodos Eigenface, Fisherface, SIFT, Mel, Mellin, Modelo Proposto com DPA. A variável P representa a precisão e a variável R a revocação.

Banco de Dados	Eigenface		Fisherface		SIFT		Mel		Mellin		Modelo Proposto	
	P	R	P	R	P	R	P	R	P	R	P	R
JAFFE	0,99	0,98	0,97	0,97	0,99	0,99	0,99	0,99	0,98	0,98	<b>1,00</b>	<b>1,00</b>
AT&T	0,94	0,94	0,89	0,82	0,92	0,90	0,84	0,81	0,83	0,81	<b>0,97</b>	<b>0,96</b>
UMIST	<b>0,99</b>	<b>0,98</b>	0,97	0,97	0,98	0,97	0,92	0,91	0,93	0,93	0,98	0,98
Georgia	0,98	0,97	0,78	0,77	0,96	0,96	0,97	0,96	0,93	0,91	<b>0,98</b>	<b>0,98</b>

Georgia, AT&T. Observando que o modelo proposto apresenta uma maior eficiência com o banco de dados AT&T em relação aos métodos Fisherface, SIFT, Mel-cepstrum e Mellin-cepstrum. Na banco de dados UMIST o melhor resultado foi obtido pelo método Eigenface como esperado. O método Fisherface obteve o pior resultado para a base Georgia devido ao fato do método considerar todas as informações da imagem e o banco de dados ter como característica um fundo confuso. Assim, o método pode se basear em características do fundo e não da face. Com base nos resultados aqui apresentados, pode-se concluir que o modelo proposto também é eficiente em outras condições com outras propriedades diferentes da variação de iluminação.

**Tabela 5.8:** Resultados dos métodos Eigenface, Fisherface, SIFT, Mel-cepstrum, Mellin-cepstrum e o Modelo Proposto. A variável *Acc* representa a acurácia e a variável *Var* a variância.

Banco de Dados	Eigenface		Fisherface		SIFT		Mel		Mellin		Modelo Proposto	
	Acc	Var	Acc	Var	Acc	Var	Acc	Var	Acc	Var	Acc	Var
JAFFE	98,65	1,70	97,69	2,07	99,23	0,51	99,23	0,51	98,46	0,22	<b>100,00</b>	0,00
AT&T	94,10	5,94	82,00	12,80	90,70	2,06	81,70	2,96	81,00	1,8	<b>96,10</b>	1,49
UMIST	<b>98,72</b>	<b>0,10</b>	97,96	0,29	98,04	0,43	93,38	1,73	94,46	0,84	98,10	0,66
Georgia	97,77	0,89	77,14	504,5	96,05	0,69	96,51	0,43	91,54	0,57	<b>97,86</b>	1,48

Para uma outra forma de analisar os resultados são apresentados os gráficos mostrados nas Figuras 5.6, 5.7, 5.8, 5.9. O primeiro gráfico é referente ao banco de dados

JAFFE, onde o modelo proposto apresentou uma melhor eficiência em relação aos outros modelos. O segundo gráfico é em relação a banco de dados AT&T e como pode ser observado o método proposto possui também melhores resultados. Já o terceiro gráfico, é referente a base UMIST onde os melhores resultados foi do método Eigenface que por ser da abordagem holística é um método que exige uma melhor estrutura computacional que o modelo proposto. Por fim, o quarto gráfico que está relacionado o banco de dados Georgia onde o modelo proposto mostra melhores resultados em relação a todos os outros métodos.

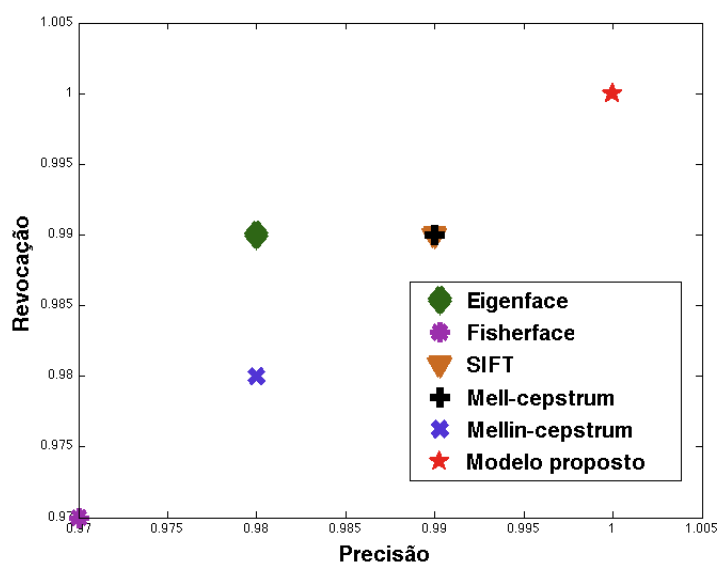


Figura 5.6: Resultado comparativo para a banco de dados JAFFE

## 5.6 Considerações Finais

Para uma melhor análise do modelo proposto foram utilizados dois grupos de bancos de dados e feitos três experimentos. O primeiro grupo de bancos de dados contém imagens com variação de iluminação, ou seja, com partes escurecidas. Este grupo é para testar a eficácia do tratamento de variação de iluminação. O segundo grupo é de bancos de dados de imagens com diferentes características tais como variação de rotação e fundo variável.

Os experimentos foram baseados na avaliação da precisão, revocação, acurácia e variância. O primeiro experimento analisa os resultados das variações do modelo pro-

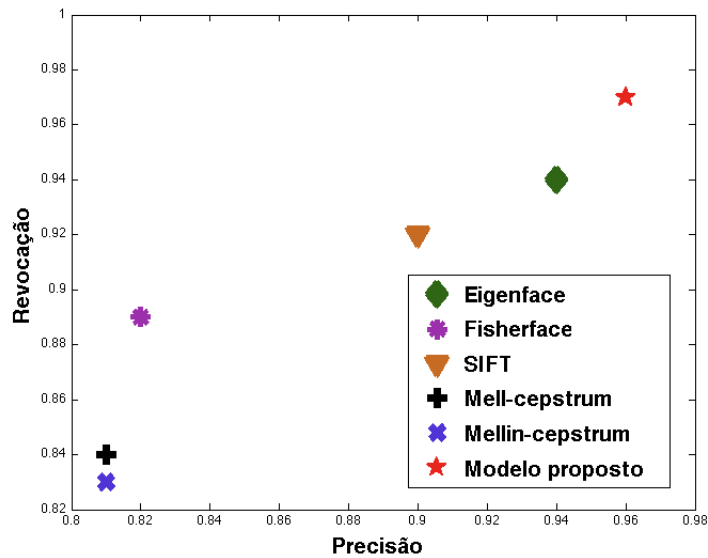


Figura 5.7: Resultado comparativo para a banco de dados AT&T

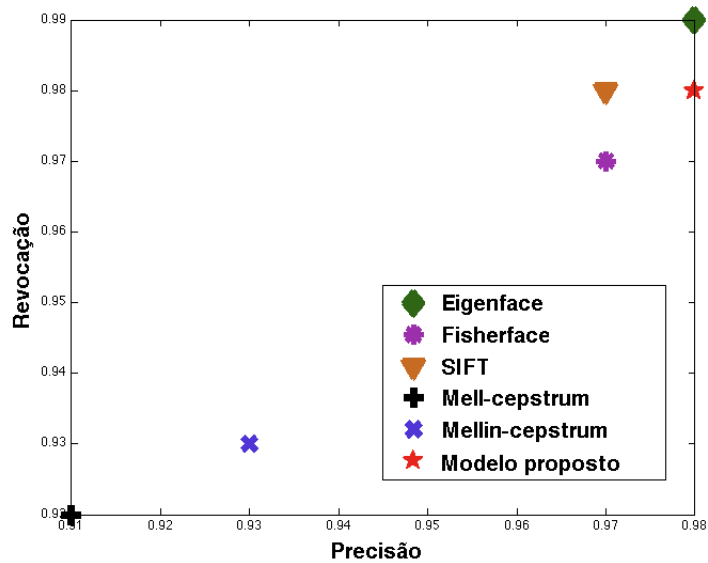
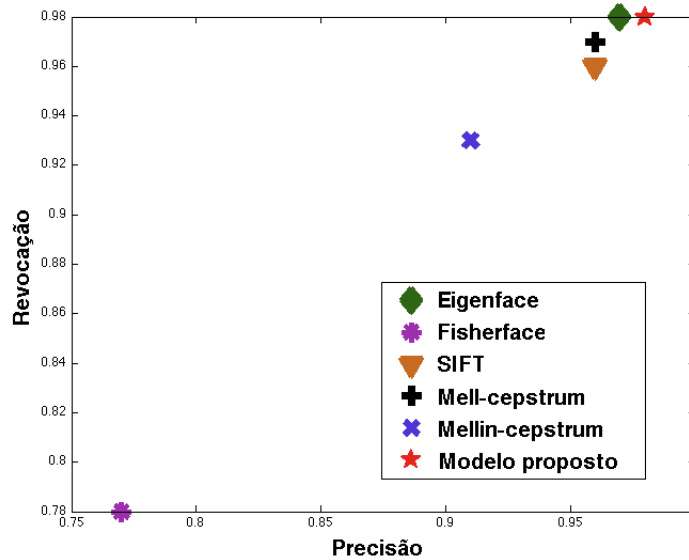


Figura 5.8: Resultado comparativo para a banco de dados UMIST

posto e como é mostrado neste capítulo o modelo com o uso do DCT para o pré-processamento obtém melhores resultados e com pouca diferença em relação ao uso ou não da redução de dimensionalidade. O segundo é feito uma análise comparativa com os métodos apresentados por (Kao, Hsu and Yang 2010) e (Chen, Meng and Shingian 2006) nos quais o modelo proposto foi baseado. E no experimento três a análise comparativa



**Figura 5.9:** Resultado comparativo para a banco de dados Georgia

é com outros métodos apresentados na literatura. Em ambos experimentos o modelo proposto apresenta melhores resultados.

No capítulo a seguir são mostradas as conclusões obtidas no trabalho. São também apresentadas as vantagens e desvantagens do modelo proposto.



# Capítulo 6

## Conclusão

O reconhecimento de face não é uma tarefa trivial. Existem muitos fatores que devem ser levados em conta na hora do reconhecimento, principalmente no mundo real. Devido a isso, ainda existe um vasto campo de estudos voltados para suprir os fatores que interferem no trabalho de reconhecimento.

A variação de iluminação está entre os fatores que mais afetam a performance dos métodos de reconhecimento de face. Isso acontece devido as sombras que são criadas na imagem evitando a análise de características relevantes utilizadas na hora do reconhecimento.

O presente trabalho propõe um método baseado na abordagem DCT que seja invariante quando há variações de iluminação. O objetivo principal é propor um método que seja estável mesmo em imagens de visão parcial devido as sombras criadas pela iluminação instável.

O modelo proposto neste trabalho possui duas versões. A primeira delas utiliza o DCT sem a redução de dimensionalidade. Como é apresentado na seção de resultados, a primeira versão apresenta uma melhor performance no reconhecimento de faces quando comparado com os outros métodos também desenvolvidos para tratar a questão de variação de iluminação. Isso também acontece quando o método, é comparado com outras abordagens apresentadas na literatura e testado em base de dados variadas. Esta primeira versão, por não fazer a redução de dimensionalidade dos vetores de características, possui uma maior quantidade de dados como entrada do processo de reconhecimento, o que supõem exigir mais da estrutura computacional quando comparada com a segunda versão que reduz os dados de entrada.

A segunda versão do método proposto segue os mesmos passos da primeira, porém, antes de realizar a tarefa de reconhecimento propriamente dita, executa a técnica DPA que seleciona as características para a classificação. Como é apresentado no capítulo de resultados, esta versão possui a mesma eficácia ou melhor quando comparada com a primeira versão. Conseqüentemente, possui melhor performance tanto quando comparado com os outros métodos desenvolvidos para lidar com a variação de iluminação, quanto os métodos selecionados na literatura desenvolvidos para lidar com diferentes propriedades tais como variação de expressão, posição, gestos e fundo.

Devido ao fato do método com a técnica DPA fazer a redução de dimensionalidade, ou seja, selecionar as características coletadas da face antes de realizar o reconhecimento, o mesmo possui um menor custo por lidar com vetores de características menores. Como ele apresenta a mesma eficiência ou até mesmo melhor, esta versão se torna mais viável para a tarefa de reconhecimento.

Analisando a performance de todos os métodos, pode ser observado que o método proposto obtém melhores resultados com um menor custo computacional por gerar menores vetores de características. Porém para a base de dados UMIST o método de melhor eficiência foi o apresentado por (Kao, Hsu and Yang 2010).

Os métodos Eigenface e Fisherface possuem uma boa performance para o reconhecimento de faces. Porém ambos pertencem à abordagem Holística. A mesma possui como propriedade, considerar todas as características da face na hora de realizar a tarefa de reconhecimento. Esta propriedade faz com que os métodos Eigenface e Fisherface trabalhe com um maior volume de dados que o método proposto que foi baseado na abordagem Extração de característica, onde são selecionadas as propriedades mais relevantes da face para realização do reconhecimento de face. Outro ponto a ser observado é que o método Fisherface foi o mais instável entre todos os métodos mostrado neste trabalho.

A desvantagem apresentada pelo modelo proposto é que depois de normalizar a imagem no domínio DCT é necessário calcular a transformada inversa, para a partir da imagem normalizada extrair novamente os coeficientes DCT, agora, em blocos. Outro ponto é o fato da taxa de compressão do DPA ser definida empiricamente.

Como trabalhos futuros o modelo proposto possui a questão de análise do valor do DPA para verificar se existe um número em que o método se mostre mais eficiente. Existe também a questão do tamanho da janela usada no LCE para denotar a luminosidade média dos *pixels* vizinhos, onde pode-se fazer uma análise do comportamento do sistema quando esse tamanho é alterado. Outro ponto é tentar melhorar o método eliminando

a necessidade de calcular a transformada inversa, para que os coeficientes DCT, possam ser extraídos diretamente da imagem normalizada no domínio DCT.



# Referências Bibliográficas

- Adini, Y., Moses, Y. and Ullman, S.: n.d., Face recognition: the problem of compensating for changes in illumination direction, *IEEE Trans. Pattern Analysis and Machine Intelligence*, year = 1997, volume = 19(7), pages = 721-732, .
- Bartlett, M. S., Movellan, J. R. and Sejnowski, T. J.: 2002, Face recognition by independent component analysis, *IEEE Trans. on Neural Networks* **13**(6), 1450–1464.
- Basri, R. and Jacobs, D. W.: 2003, Lambertian reflectance and linear sub-spaces, *IEEE Trans. Pattern Analysis and Machine Intelligence* **25**(2), 218–233.
- Batista, N. C.: 2009, Vocabulários visuais aplicados à detecção de edifícios em fotografias históricas, *Dissertação apresentada ao Instituto de Ciências Exatas da Universidade Federal de Minas Gerais* .
- Belhumeur, P., Hespanha, J. and Kriegman, D.: 1997a, Eigenfaces vs fisherfaces recognition using class specific linear projection, *IEEE Trans.PAMI* pp. 711–720.
- Belhumeur, P. and Kriegman, D.: 1998, What is the set of images of an object under all possible illumination conditions, *IJCV* **28**(3), 245–260.
- Belhumeur, P. N., Hespanha, J. P. and Kriegman, D. J.: 1997b, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **19**, 711–720.
- Belkin, M. and Niyogi, P.: 2001, Laplacian eigenmaps and spectral techniques for embedding and clustering, *Advances in Neural Information Processing System* **14**, 585–591.
- Berry, M. J. A. and Linoff, G.: 1997, Data mining techniques for marketing, sales, and customer support.

- Blanz, V. and Vetter, T.: 2003, Face recognition based on fitting a 3d morphable model, *IEEE Trans. PAMI* **25**(9), 1063–1073.
- Bowyer, K. W., Chang, K. and Flynn, P.: 2004, A survey of approaches to three-dimensional face recognition, *Proc. ICPR* .
- Bowyer, K. W., Chang, K. and Flynn, P.: 2006, A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition, *CVIU* .
- Brunelli, R. and Poggio, T.: 1993, Face recognition: features versus templates, *IEEE Trans. on Pattern Analysis and Machine Intelligence. PAMI* **15**, 1042–1052.
- Burdick, D., Calimlim, M. and Gehrke, J.: 2001, Mafia: a maximal frequent itemset algorithm for transactional databases, *Proceedings of international Conference on Data Engineering* pp. 443–452.
- Burges, C. J.: 1998a, A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery* **2**, 121–267.
- Burges, C. J. C.: 1998b, A tutorial on support vector machines for pattern recognition, *Knowledge Discovery and Data Mining* **2**(2), 1–43.
- Cakir, S. and Cetin, A. E.: 2011, Mel-and mellin-cepstral feature extraction algorithms for face recognition, *Comput. J.* **54**, 1526–1534.
- Caleanu, C. D.: 2011, Combined pattern search optimization of feature extraction and classification parameters in facial recognition, *Pattern Recognition* **32**, 1250–1255.
- Cambrige, L.: 2002, Database of faces. <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.
- Caselles, V., Lisani, J. L. and Sapiro, G.: 1999, Shape preserving local histogram modification, *IEEE Trans. Image Process.* **8** (2), 220–230.
- Cavalcanti, G. D. C. and Filho, E. C. B. C.: 2003, Eigenbands fusion for frontal face recognition, *Proceedings of IEEE International Conference on Image Processing* **1**, 665–668.
- Chang, C. and Lin, C.: 2001, Libsvm: a library for support vector machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Chen, H., Belhumeur, P. and Jacobs, D.: 2000, In search of illumination invariants, *Proc. IEEE conf. CVPR* .

- Chen, L., Liao, H., Ko, M., J., L. and Yu, G.: 2000, A new lda-based face recognition system which can solve the small samples size problem, *Journal of Pattern Recognition* **33**, 1713–1726.
- Chen, W., Meng, J. E. and Shingian, W.: 2006, Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain, *IEEE Trans. on In Systems, Man and Cybernetics, Part B* **36(2)**, 458–466.
- Chen, X., Flynn, P. J. and Bowyer, K. W.: 2003, Visible-light and infrared face recognition, *Proc. Workshop on Multimodal User Authen.*
- Colombo, C., Bimbo, A. D. and Magistris, S. D.: 1995, Human-computer interaction based on eye movement tracking, *Computer Architectures for Machine Perception* pp. 258–263.
- Cox, I. J., Ghoson, J. and Yianilos, P. N.: 1996, Feature based face recognition using mixture-distance, *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition* pp. 1–24.
- Csurka, G., Dance, C. R., Fan, L., Willamowski, J. and Bray, C.: 2004, Visual categorization with bags of keypoints.
- Dabbaghchian, S., Ghaemmaghmi, M. P. and Aghagolzadeh, A.: 2010, Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology, *Pattern Recognition* **43**, 1431–1440.
- Dai, G. and Zhou, C.: 2003, Face recognition using support vector machines with the robust feature, *Proceedings of IEEE Workshop on Robot and Human Interactive Communication* pp. 49–53.
- DeMers, D. and Cottrell, G. W.: 1993, Non-linear dimensionality reduction, *Advances in Neural Information Processing Systems* **5**, 580–587.
- D'eniz, O., Castrill'on, M. and Hern'andez, M.: 2003, Face recognition using independent component analysis and support vector machines, *Pattern Recognition Letters* **24**, 2153–2157.
- Draper, B., Baek, B. K., Bartlett, M. S. and Beveridge, J. R.: 2003, Recognizing faces with pca and ica, *Computer Vision and Image Understanding: Special Issue on Face Recognition* **91**, 115–137.

- Eleyan, A. and Demirel, H.: 2005, Face recognition system based on pca and feedforward neural networks, *Computational Intelligence and Bioinspired Systems. Lecture Notes in Computer Science: Springer Berlin / Heidelberg* **3512**, 935–942.
- Elmarsri, R. and Navathe, S. R.: 2005, *Sistemas de banco de dados*, Pearson Addison Wesley, 4<sup>th</sup> ed., São Paulo.
- Foon, N. H., Jin, A. T. B. and Ling, D. N. C.: 2004, Face recognition using wavelet transform and non-negative matrix factorization, *Advances In Artificial Intelligence, Proceedings. Lecture Notes In Artificial Intelligence* **3339**, 192–202.
- Freund, Y. and Schapire, R. E.: 1997, A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences* **55**, 119–139.
- Friedman, J. H.: 1989, Regularized discriminant analysis, *Journal of the American Statistical Association* **84**, 165–175.
- Fukunaga, K.: 1990, Introduction to statistical pattern recognition, *second ed. Boston, Ma: Academic Press* .
- Gao, W., Cao, B., Shan, S., Zhou, D., Zhang, X. and Zhao, D.: 2004, The cas-peal large-scale chinese face database and evaluation protocols, technique report, joint research & development laboratory, *The Chinese Academy of Sciences, Beijing* .
- Gao, Y. and Leung, M.: 2002, Face recognition using line edge map, *IEEE Trans. PAMI* .
- Georghiades, A., Belhumeur, P. and Kriegman, D.: 2001a, From few to many: Illumination cone models for face recognition under variable lighting and pose, *IEEE Trans. Pattern Analysis and Machine Intelligencence* **23**(6), 643–660.
- Georghiades, A. S., Belhumeur, P. N. and Kriegman, D. J.: 2001b, From few to many: illumination cone models for face recognition under variable lighting and pose, *IEEE Trans. Pattern Analysis and Machine Intelligenc* **23**(6), 643–660.
- Georgia Tech Face Database*: 2007. [http://www.anefiam.com/research/face\\_reco.html](http://www.anefiam.com/research/face_reco.html).
- Gonzalez, R. C. and Woods, R. E.: 1992, Digital image processing, *Reading, MA: Addison-Wesley* .



- Graf, H. P., Chen, T., Petajan, E. and Cosatto, E.: 1995, Locating faces and facial parts, *Internacional Workshop on Automatic Face and Gesture Recognition* pp. 41–46.
- Graham, D. B. and Allinson, N. M.: 1998, The umist database. <http://www.face-rec.org/databases/>.
- Gross, R. and Brajovic, V.: 2003, An image preprocessing algorithm for illumination invariant face recognition, *Lect. Notes Comput. Sci.* **26888**, 10–18.
- Gueham, M., Bouridane, A., Crookes, D. and Nibouche, O.: 2008, Automatic recognition of shoeprints using fourier mellin transform, *NASA/ESA Conference on Adaptive Hardware and Systems* pp. 487–491.
- Guo, G., Li, S. and Kapluc, C.: 2000, Face recognition by support vector machines, *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition. Washington, DC, USA* pp. 196–201.
- Hallinan, P.: 1994, A low-dimensional representation of human faces for arbitrary lighting conditions, *Proc. IEEE conf. CVPR* .
- Han, J. and Kamber, M.: 2001, Data mining - concepts and techniques, *Morgan Kaufmann* .
- Hastie, T. and Tibshirani, R.: 1998, Classification by pairwise coupling, *The Annals of Statistics* **26**, 451–471.
- He, X., Yan, S. C., Hu, Y. X. and Zhang, H. J.: 2003, Learning a locality preserving subspace for visual recognition, *Proceedings of 9th IEEE International Conference on Computer Vision* **1**, 385–392.
- He, X., Yan, S., Hu, Y., Niyogi, P. and Zhang, H.-J.: 2005, Face recognition using laplacianfaces, *IEEE Trans. PAMI* **27(3)**, 328–340.
- Horn, B. K. P.: 1986, Robot vision, *Cambridge, MA: MIT Press* .
- Hotelling, H.: 1933, Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology* **24**, 417–441, 498–520.
- Howland, P. and Park, H.: 2004, Generalized discriminant analysis using the generalized singular value decomposition, *IEEE Trans. On Pattern Analysis and Machine Intelligence* **26**, 995–1006.

- Hsu, C. W., Chang, C. C. and Lin, C. J.: 2009, A practical guide to support vector classification, *Taipei, Taiwan. Acessado em 22 de junho de 2009* . <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- Hu, H.: 2008a, Ica-based neighborhood preserving analysis for face recognition, *Computer Vision and Image Understanding* **112**, 286–295.
- Hu, H.: 2008b, Orthogonal neighborhood preserving discriminant analysis for face recognition, *Pattern Recognition* **41**, 2045–2054.
- Huang, R. J. J.: 1998, Detection strategies for face recognition using learning and evolution, *George Mason University, Fairfax, Virginia, PhD. Dissertation* .
- Ishiyama, R., Hamanaka, M. and Sakamoto, S.: 2005, An appearance model constructed on 3-d surface for robust face recognition against pose and illumination variations, *IEEE Trans. Syst. Man Cybern. C* **35**(3), 326–334.
- Jafri, R. and Arabnia, H. R.: 2009, A survey of face recognition techniques, *Journal of Information Preccessing System* **5**(2), 41–68.
- Jain, A. K. and Dubes, R. C.: 1988, Algorithms for clustering data, *New Jersey: Prentice-Hall* .
- Jiang, Y. G., Ngo, C.-W. and Yang, J.: 2007, Towards optimal bag-of-features for object categorization and semantic video retrieval, *Proceedings of the ACM International Conference on Image and Video Retrieval* pp. 494–501 New York, NY, USA. ACM.
- Jing, X. Y. and Zhang, D.: 2004, A face and palmprint recognition approach based on discriminant dct feature extraction, *IEEE Trans. on Systems, Man and Cybernetics* **34**(6), 2405–2415.
- Jobson, D. J., Rahman, Z. and Woodel, G. A.: 1996, Properties and performance of a center/surround retinex, *IEEE Trans. on Image Processing: special issue on color processing* .
- Jobson, D. J., Rahman, Z. and Woodel, G. A.: 1997, Amultiscaleretinexfor bridging the gap between color images and the human observation of scences, *IEEE Trans. on Image Processing* **6**(7).
- K. C. Chang, K. C., Bowyer, K. W. and Flynn, P. J.: 2005, An evaluation of multimodal 2d+3d face biometrics, *IEEE Trans. PAMI* **27**(4).

- K., S. L. and T., R. S.: 2003, Think globally, fit locally: Unsupervised learning of low dimensional manifolds, *Machine Learning Research* **4**, 119–155.
- Kamachi, M., Lyons, M. and Gyoba, J.: 1998, The japanese female facial expression (jaffe) database. <http://www.kasrl.org/jaffe.html>.
- Kanade, T.: 1973, Picture processing system by computer complex and recognition of human faces, *Kyoto University, Japan, PhD. Thesis* .
- Kao, W., Hsu, M. and Yang, Y.: 2010, Local contrast enhancement and adaptive feature extraction for illumination-invariant face recognition, *Pattern Recognition* **43**(5), 1736–1747.
- Kim, J.Y. and Kim, L. and Hwang, S.: 2001, An advanced contrast enhancement using partially overlapped sub-block histogram equalization, *IEEE Trans. Circuits Syst. Video Technol.* **11** (4), 475–484.
- Kittler, J., Hilton, A., Hamouz, M. and Illingworth, J.: 2005, 3d assisted face recognition: A survey of 3d imaging, modelling and recognition approaches, *Proc. IEEE conf CVPR* .
- Kong, S., Heo, J., Abidi, B., Paik, J. and Abidi, M.: 2004, Recent advances in visual and infrared face recognition - a review, *CVIU* .
- Krebel, U.: 1999, Pairwise classification and support vector machines, *Advance in Kernel Methods - Support Vector Learning* pp. 255–268.
- Kwak, K. C. and Pedrycz, W.: 2007, Face recognition using an enhanced independent component analysis approach, *IEEE Trans. on Neural Networks* **18**, 530–541.
- Kyperountas, M., Tefas, A. and Pitas, I.: 2010, Salient feature and reliable classifier selection for facial expression classification, *Pattern Recognition* **43**, 972–986.
- Lee, D. D. and Seung, H. S.: 1999, Learning the parts of objects by non-negative matrix factorization, *Nature* **401**, 788–791.
- Lee, J. K., Kabrisky, M., Oxley, M. E., Rogers, S. K. and Ruck, D. W.: 1993, The complex cepstrum applied to two-dimensional images, *Pattern Recognit* **26**, 1579–1592.

- Lee, K. C., Ho, J. and Kriegman, D. J.: 2005, Acquiring linear subspaces for face recognition under variable lighting, *IEEE Trans. Trans. Pattern Analysis and Machine Intelligenc* **27(5)**, 684–698.
- Li, B. and Yin, H.: 2005, Face recognition using rbf neural networks and wavelet transform, *Advances in Neural Networks - ISNN 2005* **3497**, 105–111.
- Li, H. Q., Wang, S. Y. and Qi, F. H.: 2004, Automatic face recognition by support vector machines, *Combinatorial Image Analysis, Proceedings* **3322**, 716–725.
- Li, J., Zhou, S. and Shekhar, C.: 2003, A comparison of subspace analysis for face recognition, *Proc. IEEE Intl Conf. on Acoustics, Speech, and Signal Processing* pp. 121–124.
- Liang, Y., Gong, W., Pan, Y., Li, W. and Hu, Z.: 2005, Gabor features-based classification using svm for face recognition, *Advances in Neural Networks ISNN* **3497**, 118–123.
- Liu, C.: 2004, Enhanced independent component analysis and its application to content based face image retrieval, *IEEE Trans. Syst. Man Cybern.* **34(2)**, 1117–1127.
- Liu, S. and Ruan, Q.: 2011, Orthogonal tensor neighborhood preserving embedding for facial expression recognition, *Pattern Recognition* **44**, 1497–1513.
- Liu, W., Wang, Y., Li, S. Z. and Tan, T.: 2004a, Nearest intra-class space classifier for face recognition, *The 17th International Conference on Pattern Recognition (ICPR) 4. Cambridge, UK*, 495–498.
- Liu, W., Wang, Y., Li, S. Z. and Tan, T.: 2004b, Null space approach of fisher discriminant analysis for face recognition, *Biometric Authentication, Lecture Notes in Computer Science: Springer Berlin / Heidelberg* **3087**, 32–44.
- Loog, M., Duin, R. P. W. and Haeb-Umbach, R.: 2001, Multiclass linear dimension reduction by weighted pairwise fisher criteria, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **23**, 762–766.
- Lorena, A. C. and Carvalho, A. C. P. L. F.: 2007, Uma introdução às support vector machines, *RITA* **14(2)**.
- Lowe, D. G.: 2004, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vision* **60**, 91–110.

- Lu, J. W., Plataniotis, K. N. and Venetsanopoulos, A. N.: 2003, Face recognition using lda-based algorithms, *IEEE Trans. On Neural Networks* **14**, 195–200.
- Lu, J. W., Plataniotis, K. N. and Venetsanopoulos, A. N.: 2005, Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition, *Pattern Recognit.* **26(2)**, 181–191.
- Martínez, A. M. and Kak, A. C.: 2001, Pca versus lda, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **23**, 228–233.
- Melin, P., Felix, C. and Castillo, O.: 2005, Face recognition using modular neural networks and the fuzzy sugeno integral for response integration, *International Journal Of Intelligent Systems* **20**, 275–291.
- Meng, J. and Zhang, W.: 2007, Volume measure in 2d pca based face recognition, *Pattern Recognition Letters* **28**, 1203–1208.
- MIT-CMU: n.d., Face dataset. [http://vasc.ri.cmu.edu/idb/html/face/frontal\\_images](http://vasc.ri.cmu.edu/idb/html/face/frontal_images).
- Moghaddam, B. Nastar, C. and Pentland, A.: 1996, A bayesian similarity measure for direct image matching, *Proceedings 13th International Conference on Pattern Recognition* pp. 350–358.
- Montgomery, D. C. and Runger, G. C.: 2003, *Applied statistic and probability for engineers*, John Wiley & Sons, 3rd ed., United States.
- Moreira, M. and Mayoraz, E.: 1998, Improved pairwise coupling classification with correcting classifiers, *Proceedings of the 10th European Conference on Machine Learning* **1398, Lecture Notes In Computer Science. London, UK: Springer-Verlag**, 160–171.
- Nixon, M.: 1985, Eye spacing measurement for facial recognition, *SPIE Proceedings* pp. 279–285.
- Nowak, E., Jurie, F. and Triggs, B.: 2006, Sampling strategies for bag-of-features image classification, *Proceedings of the European Conference on Computer Vision* pp. 490–503.
- Ojala, T., Pietikainen, M. and Maenpaa, M.: 2002, Multi-resolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **24**, 971–987.

- Oppenheim, A. V., Schaffer, R. W. and Buck, J. R.: 1999, Discrete-time signal processing (2nd edn), *Prentice-Hall, Upper Saddle River, NJ, USA* .
- Pan, Z., Rust, A. G. and Bolouri, H.: 2000, Image redundancy reduction for neural network classification using discrete cosine transforms, *Proceeding of International Joint Conference on Neural Networks (Como, Italy)* **3**, 149–154.
- Pearson, K.: 1901, On lines and planes of closest fit to systems of points, *Philosophical Magazine* **6**(2), 559–572.
- Pentland, A., Moghaddam, B. and Starner, T.: 1994, Viewbased and modular eigenspaces for face recognition, *IEEE Conference on Computer Vision and Pattern Recognition* pp. 84–90.
- Phillips, P. J., Wechsler, H., Huang, J. and Rauss, P.: 1998, The feret database and evaluation procedure for face recognition algorithms, *Image Vis. Comput* **16**(5).
- Reisfeld, D.: 1994, Generalized symmetry transforms: attentional mechanisms and face recognition, *Tel-Aviv University, PhD. Thesis, technical report* .
- Roeder, N. and Li, X.: 1995, Experiments in analyzing the accuracy of facial feature detection, *Vision Interface* pp. 8–16.
- Roweis, S. T. and Saul, L. K.: 2000, Nonlinear dimensionality reduction by locally linear embedding, *Science* **290**, 2323–2326.
- Sanderson, C. and Paliwal, K. K.: 2003, Features for robust face-based identity verification, *Journal of Signal Processing* **83**, 931–940.
- Santos, A. R.: 2005, Identificação de faces humanas através de pca-lda e redes neurais som, *Dissertação apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo* .
- Savvides, M. and Kumar, V.: 2003, Illumination normalization using logarithm transforms for face authentication, *Proc. IAPR AVBPA* pp. 549–556.
- Shan, S., Gao, W., Cao, B. and Zhao, D.: 2003a, Illumination normalization for robust face recognition against varying lighting conditions, *Proc. IEE Workshop on AMFG* pp. 157–164.

- Shan, S., Gao, W., Cao, B. and Zhao, D.: 2003b, Illumination normalization for robust face recognition against varying lighting conditions, *Proc. IEEE workshop on AMFG* .
- Shashua, A.: 1997, On photometric issue in 3d visual recognition from a single 2d image, *IJCV* .
- Shashua, A. and Riklin-Raviv, T.: 2001, The quotient image: class-based re-rendering and recognition with varying illuminations, *IEEE Trans. Pattern Analysis and Machine Intelligence* **23**(2), 129–139.
- Shim, H., Luo, J. and Chen, T.: 2008, A subspace model-based approach for face relighting under unknown lighting and poses, *IEEE Trans. Image Process* **17** (8), 1331–1341.
- Short, J.: 2006, Illumination invariance for face verification, *PhD thesis, University of Surrey* .
- Sirovich, L. and Kirby, M.: 1987, Low-dimensional procedure for the characterization of human faces, *Journal of the Optical Society of American A: Optics, Image Science, and Vision* **4**, 519–524.
- Sirovitch, L. and Kirby, M.: 1987, Low-dimensional procedure for the characterization of human faces, *J. Optical Soc. of America A* **2**, 519–524.
- Socolinsky, D. and Selinger, A.: 2004, Thermal face recognition over time, *Proc. ICPR* .
- Socolinsky, D. and Selinger, A.. In, .: 2002, A comparative study of face recognition performance with visible and thermal infrared imagery, *Proc. ICPR* .
- Socolinsky, D., Selinger, A. and Neuheise, J.: 2003, Face recognition with visible and thermal infrared imagery, *CVIU* pp. 72–114.
- Song, F., Zhang, D., Wang, J., Liu, H. and Tao, Q.: 2007, A parameterized direct lda and its application to face recognition, *Neurocomputing* **71**, 191–196.
- Swets, D. L. and Weng, J. J.: 1996, Using discriminant eigenfeatures for image retrieval, *IEEE Trans. On Pattern Analysis And Machine Intelligence* **18**, 831–836.
- Tenenbaum, J., Silva, V. and Langford, J.: 2000, A global geometric framework for nonlinear dimensionality reduction, *Science* **290**, 2319–2323.

- Toreyin, B. U. and Cetin, A. E.: 2009, Shadow detection using 2d cepstrum, *In Acquisition, Tracking, Pointing, and Laser Systems Technologies XXIII, Orlando, FL, USA, 733809. SPIE* .
- Tsao, W., Lee, A. J. T., Lui, Y., Chang, T. and Lin, H.: 2010, A data mining approach to face detection, *Pattern Recognition* **43**, 1039–1049.
- Turk, M. and Pentland, A.: 1991a, Eigenfaces for recognition, *Journal of Cognitive Neuroscience* **3(1)**, 71–86.
- Turk, M. and Pentland, A.: 1991b, Face recognition using eigenface, *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition* pp. 586–591.
- Turk, M. and Pentland, A.: 1991c, Face recognition using eigenfaces, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pp. 586–591.
- Turk, M. and Pentland, A.: 1994, Visual computing: the integration of computer graphics, visual perception and imaging.
- Vasilescu, M. A. O. and Terzopoulos, D.: 2003, Multilinear subspace analysis of image ensembles, *IEEE International Conference on Computer Vision and Pattern Recognition* pp. 93–99.
- Vedaldi, A. and Fulkerson, B.: 2008, Vlfeat: An open and portable library of computer vision algorithms, <http://www.vlfeat.org/>.
- Venkataramani, K., Qidwai, S. and Vijayakumar, B. V. K.: 2005, Face authentication from cell phone camera images with illumination and temporal variations, *IEEE Trans. Syst. Man Cybern. C* **35(3)**, 411–418.
- Wang, X. and Tang, X.: 2004, Dual-space linear discriminant analysis for face recognition, *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition* pp. 564–569.
- Wang, Y., Liu, Z., Hua, G., Wen, Z., Zhang, Z. and Samaras, D.: 2007, Face re-lighting from a single image under harsh lighting conditions, *IEEE Conference on Computer Vision and Pattern Recognition* .
- Wei, S. and Lai, S.: 2004, Robust face recognition under lighting variations, *Proc. ICPR*



- Wen, Z., Liu, Z. and Huang, T. S.: 2003, Face relighting with radiance environment maps, *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on* **2**, 158.
- Wilder, J., Phillips, P. J., Jiang, C. and Wiener, S.: 1996, Comparison of visible and infrared imagery for face recognition, *Proc. IEEE AFGR* .
- Wu, Y., Chan, K. L. and Wang, L.: 2004, Face recognition based on discriminative manifold learning, *Proc. IEEE International Conference on Pattern Recognition* **4**, 171–174.
- Xie, X. and Lam, K. M.: 2008, Face recognition using elastic local reconstruction based on a single face image, *Pattern Recognition* **41**, 406–417.
- Yan, S. C., Zhang, H. J., Hu, Y. X., Zhang, B. Y. and Cheng, Q. S.: 2001, Discriminant analysis on embedded manifold, *European Conference on Computer Vision. Springer Berlin / Heidelberg LNCS* **3021**, 121–132.
- Yang, C., Lai, S. and Chang, L.: 2004, Robust face matching under different lighting conditions, *URASIP Journal on App. Sig. Proc.* .
- Yang, J. and Zhang, D.: 2004, Two-dimensional pca: A new approach to appearance-based face representation and recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence* **28**, 131–137.
- Yang, Q. and Ding, X. Q.: 2003, Symmetrical principal component analysis and its application in face recognition, *Chinese Journal of Computers* **26**, 1146–1151.
- Yang, Q. and Tang, X.: 2004, Recent advances in subspace analysis for face recognition, *SINOBIOMETRICS* pp. 275–287.
- Yang, W., Sun, C. and Zhang, L.: 2011, A multi-manifold discriminant analysis method for image feature extraction, *Pattern Recognition* **44**, 1649–1657.
- Ye, J. P., Janardan, R., Park, C. H. and Park, H.: 2004, An optimization criterion for generalized discriminant analysis on undersampled problems, *IEEE Trans. On Pattern Analysis and Machine Intelligence* **26**, 982–994.
- Yeshurun, Y. and Schwartz, E.: n.d., Cepstral filtering on a columnar image architecture: a fast algorithm for binocular stereo segmentation, *IEEE Trans. Pattern Analysis and Machine Intelligenc*, volume = 11, year = 1989, pages = 759-767, .

- Yu, H. and Yang, J.: 2001, A direct lda algorithm for high-dimensional data with application to face recognition, *Pattern Recognition* **34**, 2067–2070.
- Yuille, A., Cohen, D. and Hallinan, P.: 1998, Feature extraction from faces using deformable templates, *IEEE Computer Society Conference on Computer Vision and Templates. San Diego, CA, USA* pp. 104–109.
- Zhang, G. C., Huang, X. S., Li, S. Z., Wang, Y. S. and Wu, X. H.: 2004, Boosting local binary pattern (lbp)-based face recognition, *Advances In Biometric Person Authentication, Proceedings* **3338**, 179–186.
- Zhang, J., Li, S. Z. and Wang, J.: 2004, Nearest manifold approach for face recognition, *Proc. IEEE International Conference on Automatic Face and Gesture Recognition* pp. 223–228.
- Zhao, J., Su, Y., Wang, D. and Luo, S.: 2003, Illumination ratio image: synthesizing and recognition with varying illuminations, *Pattern Recognit. Lett.* **24**, 2703–2710.
- Zhao, L. and Samarasinghe, D.: 2003, Face recognition under variable lighting using harmonic image exemplars, *Proc. IEEE Conf. Computer Vision and Pattern Recognition* **1**, 19–25.
- Zhou, D. and Yang, X.: 2004, Face recognition using direct-weighted lda, *8th Pacific Rim International Conference on Artificial Intelligence. Auckland, New Zealand* pp. 760–768.
- Zou, X., Kittler, J. and Messer, K.: 2007, Illumination invariant face recognition: A survey, *IEEE*.