

**UNIVERSIDADE FEDERAL DO PARANÁ**

**JHONY KAESEMODEL PONTES**

**ESTIMATIVA HIERÁRQUICA DA IDADE BASEADA EM  
CARACTERÍSTICAS GLOBAIS E LOCAIS DE IMAGENS  
FACIAIS**

**CURITIBA**

**2015**

**JHONY KAESEMODEL PONTES**

**ESTIMATIVA HIERÁRQUICA DA IDADE BASEADA EM  
CARACTERÍSTICAS GLOBAIS E LOCAIS DE IMAGENS  
FACIAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica, Área de Concentração Sistemas Eletrônicos, Departamento de Engenharia Elétrica, Setor de Tecnologia, Universidade Federal do Paraná, como parte das exigências para a obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Alessandro L. Koerich

Coorientador: Prof. Dr. Clinton Fookes

**CURITIBA**

**2015**

## Agradecimentos

Ao grande professor e cientista Alessandro Koerich pela orientação, confiança e exemplo científico que representa.

*To Clinton Fookes for his constant support, encouragement and insightful suggestions. And also, thanks for believing in me.*

Aos professores e cientistas Alceu de Souza Britto Junior, Alessandro Zimmer e Lucas Ferrari pelas valiosas contribuições dadas.

Ao amigo e cientista Marcos Hara por servir de inspiração para esta jornada e para a vida.

A minha esposa e companheira Aline Freire por sua paciência, compreensão e incentivo nas infinitas horas de estudo necessárias para a realização deste trabalho.

Aos meus pais, Ivonete e Audir, cujos sacrifícios ao longo de muitos anos plantaram-me o amor pelo conhecimento, ciência e sabedoria.

E a todos aqueles que diretamente ou indiretamente contribuíram positivamente para que essa jornada fosse possível.

## Epígrafe

*"The known is finite, the unknown infinite; intellectually we stand on an islet in the midst of an illimitable ocean of inexplicability. Our business in every generation is to reclaim a little more land, to add something to the extent and the solidity of our possessions."*

Thomas Henry Huxley (1825-1895).



# Sumário

<b>Lista de Figuras</b> .....	<b>ix</b>
<b>Lista de Tabelas</b> .....	<b>xiv</b>
<b>Lista de Siglas</b> .....	<b>xvi</b>
<b>Resumo</b> .....	<b>xvii</b>
<b>Abstract</b> .....	<b>xviii</b>
<b>1 Introdução</b> .....	<b>1</b>
1.1 Definição do problema .....	2
1.2 Hipóteses de pesquisa .....	3
1.3 Objetivos .....	4
1.3.1 Objetivo geral .....	4
1.3.2 Objetivos específicos .....	4
1.4 Justificativa.....	4
1.5 Contribuições .....	5
1.6 Organização .....	6
<b>2 Revisão bibliográfica</b> .....	<b>7</b>
2.1 Representação do envelhecimento facial .....	8
2.1.1 Modelos antropométricos .....	8
2.1.2 Modelos de aparência ativa.....	11
2.1.3 Subespaço do padrão de envelhecimento .....	11

2.1.4	<i>Age manifold</i> .....	13
2.1.5	Modelos de aparência .....	15
2.2	Algoritmos para estimação etária .....	16
2.2.1	Classificação .....	16
2.2.2	Regressão .....	17
2.2.3	Abordagens híbridas .....	17
2.3	Análise e síntese .....	18
<b>3</b>	<b>Fundamentação teórica</b> .....	<b>22</b>
3.1	Um breve estudo sobre envelhecimento .....	22
3.2	Extração de características .....	24
3.2.1	Modelos de aparência ativa .....	24
3.2.1.1	Modelagem estatística .....	24
3.2.1.2	Modelo estatístico da forma .....	25
3.2.1.3	Modelo estatístico de textura .....	27
3.2.1.4	Modelo estatístico combinado .....	28
3.2.2	<i>Gabor wavelets</i> .....	30
3.2.3	<i>Local binary patterns</i> .....	33
3.2.4	<i>Local phase quantization</i> .....	38
3.3	Classificação e regressão .....	42
3.3.1	Máquina de vetores de suporte .....	42
3.3.1.1	Definição matemática .....	43
3.3.1.2	O truque do kernel .....	46
3.3.1.3	Margens suaves .....	47
3.3.1.4	Multiclasses .....	47
3.3.2	Máquina de vetores de suporte para regressão .....	48

<b>4</b>	<b>Método proposto</b>	<b>50</b>
4.1	Conjuntos de dados	51
4.1.1	FG-NET <i>Aging</i>	52
4.1.2	MORPH <i>Album2</i>	52
4.2	Pré-processamento	55
4.3	Regiões de interesse	57
4.4	Extração de características	61
4.4.1	Características globais	61
4.4.2	Características locais	61
4.5	Fusão das características	61
4.6	Redução do espaço de características	63
4.7	Estimativa da idade	64
<b>5</b>	<b>Resultados experimentais</b>	<b>66</b>
5.1	Extração e análise das características	67
5.1.1	Modelos de aparência ativa	67
5.1.2	Gabor <i>wavelets</i>	71
5.1.3	<i>Local binary patterns</i>	74
5.1.4	<i>Local phase quantization</i>	76
5.2	Métricas de avaliação	79
5.3	Protocolos de avaliação	80
5.4	Conjunto de dados FG-NET <i>Aging</i>	80
5.4.1	Estimativa <i>Single-level</i>	80
5.4.2	Estimativa hierárquica com margens fixas	81
5.4.2.1	Classificação em faixas etárias	81
5.4.2.2	Estimativa específica da idade	84
5.4.3	Estimativa hierárquica com regiões sobrepostas	85

5.4.4	Experimentos utilizando o Protocolo LOPO .....	90
5.5	Conjunto de dados MORPH <i>Album 2</i> .....	92
5.5.1	Estimativa hierárquica com margens fixas .....	92
5.5.1.1	Classificação em faixas etárias .....	92
5.5.1.2	Estimativa específica da idade .....	94
5.5.2	Estimativa hierárquica com regiões sobrepostas .....	95
5.6	Combinação dos conjuntos de dados FG-NET e MORPH .....	98
5.6.1	Estimativa hierárquica com margens fixas .....	99
5.6.1.1	Classificação em faixas etárias .....	99
5.6.1.2	Estimativa específica da idade .....	101
5.6.2	Estimativa hierárquica com regiões sobrepostas .....	101
5.7	Estimativa realizada por seres humanos.....	105
5.8	Comparação com outros métodos .....	108
5.8.1	FG-NET <i>Aging</i> .....	108
5.8.2	MORPH <i>Album 2</i> .....	109
<b>6</b>	<b>Conclusões.....</b>	<b>110</b>
6.1	Contribuições .....	112
6.2	Trabalhos futuros.....	113
	<b>Referências .....</b>	<b>114</b>
	<b>Apêndice A – Análise <i>local binary patterns</i> .....</b>	<b>118</b>
	<b>Apêndice B – Análise <i>local phase quantization</i> .....</b>	<b>124</b>
	<b>Apêndice C – Estimativa hierárquica por décadas .....</b>	<b>130</b>
	<b>Apêndice D – Resultados FG-NET <i>Aging</i> .....</b>	<b>133</b>

Apêndice E - Resultados MORPH <i>Album 2</i> .....	135
Apêndice F - Formulário de pesquisa.....	136

## Lista de Figuras

Figura 1	Exemplos de informações faciais [1]. . . . .	1
Figura 2	Exemplo da faixa etária de Albert Einstein [2]. . . . .	7
Figura 3	Distribuição da pressão dentro de um objeto esférico preenchido com fluido e crescimento facial simulado no perfil da cabeça de uma criança utilizando o modelo de transformação de tensão cardioidal revisado [3]. . . . .	9
Figura 4	As seis proporções de distância em imagens frontais da face [4]. . . . .	10
Figura 5	(a) Regiões da face contendo rugas faciais e (b) definição das rugas na região da testa encontradas utilizando <i>snakelets</i> [4]. . . . .	10
Figura 6	Vetorização do padrão de envelhecimento. As idades (0-8) são marcadas numa ordem temporal e o vetor de características é identificado com as imagens que estão disponíveis ( $b_i$ ) e as imagens faltantes são identificados como $m$ [5]. . . . .	13
Figura 7	<i>Aging manifold</i> utilizando PCA, OLPP e CEA [6] [7]. . . . .	15
Figura 8	Esboços de envelhecimento facial de 30 a 80 anos com 10 anos por esboço [8]. . . . .	23
Figura 9	Mudanças no contorno facial devido ao envelhecimento [8]. . . . .	24
Figura 10	Imagem facial com os pontos de referência, conjunto de pontos e o formato livre da imagem, respectivamente [9]. . . . .	25
Figura 11	Exemplo de Gabor <i>wavelet</i> [10]. . . . .	31
Figura 12	Exemplo Gabor <i>wavelet</i> 2D [10]. . . . .	32
Figura 13	Exemplo da transformada Gabor <i>wavelet</i> [10]. . . . .	32
Figura 14	Processo de construção de um código LBP [10]. . . . .	34
Figura 15	LBP invariante a translação [10]. . . . .	34
Figura 16	Padrões circulares para (P,R) [10]. . . . .	35
Figura 17	Padrões invariantes a rotação para LBP <sub>S(8,1)</sub> [10]. . . . .	36
Figura 18	Histogramas de LBP uniformes [10]. . . . .	37
Figura 19	Imagens faciais com variações na iluminação e borramentos [11]. . . . .	38
Figura 20	Exemplo de hiperplano ótimo para padrões linearmente separáveis. . . . .	44
Figura 21	Interpretação geométrica da distância algébrica. . . . .	45
Figura 22	Mapeamento não-linear $\varphi(\cdot)$ do espaço de entrada (2D) para o espaço de características (3D). . . . .	46
Figura 23	Exemplo de variáveis de folga em um SVM com margens suaves. . . . .	48
Figura 24	Exemplo de SVR linear com a margem de erro $\epsilon$ e variáveis de folga. . . . .	49

Figura 25	Método proposto para a estimativa da idade baseada na análise de imagens faciais. ....	51
Figura 26	Exemplos de imagens faciais da base de dados FG-NET. ....	52
Figura 27	Face com os 68 pontos de referência ( <i>landmarks</i> ). ....	53
Figura 28	Distribuição das amostras por idade (FG-NET). ....	54
Figura 29	Exemplos de imagens faciais da base de dados MORPH. ....	54
Figura 30	Distribuição das amostras por idade do conjunto de dados selecionado (MORPH). ....	55
Figura 31	Imagens em níveis de cinza com os 68 <i>landmarks</i> (2, 17, 28 e 69 anos respectivamente) (FG-NET). ....	56
Figura 32	Imagens redimensionadas e rotacionadas utilizando transformação de similaridades (2, 17, 28 e 69 anos respectivamente) (FG-NET). ....	56
Figura 33	Classificação de rugas faciais proposta por Lemperle [12]: (a) posição das rugas e (b) descrição. ....	57
Figura 34	As 11 regiões de interesse para a extração de características locais (2, 17, 28 e 69 anos respectivamente) (FG-NET). ....	58
Figura 35	Regiões de interesse (2 anos) (FG-NET). ....	59
Figura 36	Regiões de interesse (17 anos) (FG-NET). ....	59
Figura 37	Regiões de interesse (28 anos) (FG-NET). ....	60
Figura 38	Regiões de interesse (69 anos) (FG-NET). ....	60
Figura 39	Método para a fusão de características globais e locais. ....	64
Figura 40	Método de classificação hierárquica com regiões sobrepostas na etapa de estimação específica da idade. ....	65
Figura 41	Os 68 <i>landmarks</i> descrevendo os formatos de componentes faciais e a triangulação para a descrição da textura, respectivamente. ....	68
Figura 42	Imagem original, imagem reconstruída a partir dos parâmetros combinados do AAM e o modelo residual gerado pela diferença absoluta entre a imagem original e reconstruída. ....	69
Figura 43	Variações das 5 primeiras características combinadas do AAM em ambas as direções começando com a face média (1), sempre mostrada ao centro (FG-NET). ....	70
Figura 44	Exemplos de faces reconstruídas a partir do modelo AAM combinado (2, 17, 28 e 69 anos respectivamente) (FG-NET). ....	70
Figura 45	Filtros e magnitudes de Gabor para 8 orientações (colunas) e 5 escalas (linhas): (a) filtros e (b) magnitudes. ....	72
Figura 46	Regiões da testa utilizadas para a aplicação dos filtros de Gabor (FG-NET): (a) 2 anos e (b) 69 anos de idade. ....	72
Figura 47	Magnitudes dos filtros de Gabor e seus valores médios para 8 orientações (colunas) e 5 escalas (linhas) das regiões da testa: (a) 2 anos e (b) 69 anos de idade. ....	72

Figura 48	Regiões periorbital utilizada para a aplicação dos filtros de Gabor (FG-NET): (a) 2 anos e (b) 69 anos de idade. ....	73
Figura 49	Componentes reais dos filtros de Gabor para 8 orientações (colunas) e 5 escalas (linhas) das regiões periorbital: (a) 2 anos e (b) 69 anos de idade. ....	73
Figura 50	Diferentes padrões detectados pelo operador LBP [13]. ....	74
Figura 51	Histogramas dos códigos $LBP_{8,2}^{u2}$ para a região das linhas horizontais da testa para 2, 17, 28 e 69 anos respectivamente (FG-NET). ....	75
Figura 52	Histogramas dos códigos $LBP_{8,2}^{u2}$ para a região das linhas periorbital direitas para 2, 17, 28 e 69 anos respectivamente (FG-NET). ....	77
Figura 53	Histogramas dos códigos $LPQ_{5 \times 5}$ para a região das linhas horizontais da testa e suas respectivas orientações para 2, 17, 28 e 69 anos respectivamente (FG-NET). ....	78
Figura 54	Histogramas dos códigos $LPQ_{5 \times 5}$ para a região das linhas periorbital direita e suas respectivas orientações para 2, 17, 28 e 69 anos respectivamente (FG-NET). ....	79
Figura 55	Comparação do desempenho da classificação utilizando kernel RBF (LIBSVM) e linear (LIBLINEAR) em diferentes conjuntos de características (FG-NET). ....	82
Figura 56	Matriz de confusão da classificação em faixas etárias utilizando a combinação de características globais e locais $AAM+LPQ_{(7 \times 7)}$ (FG-NET). ..	83
Figura 57	$MAE$ de diferentes métodos de estimação etária em diferentes conjuntos de características (FG-NET). ....	85
Figura 58	Exemplo dos regressores SVR da estimativa hierárquica com regiões sobrepostas fixas em 5 anos (FG-NET). ....	86
Figura 59	Idades utilizadas para as regiões sobrepostas flexíveis entre os SVRs (FG-NET). ....	87
Figura 60	Desempenho em termos de $MAE$ dos regressores SVR da estimativa hierárquica com regiões sobrepostas flexíveis em diferentes conjuntos de características (FG-NET). ....	87
Figura 61	Escore acumulativo ( $CS$ ) dos métodos de estimativa da idade (FG-NET). ....	89
Figura 62	Escore acumulativo ( $CS$ ) do método proposto utilizando os protocolos 75%/25% e LOPO (FG-NET). ....	91
Figura 63	Desempenho da classificação utilizando kernel linear (LIBLINEAR) em diferentes conjuntos de características (MORPH). ....	93
Figura 64	Matriz de confusão da classificação em faixas etárias utilizando características locais $LPQ_{(7 \times 7)}$ (MORPH). ....	94
Figura 65	Comparação em termos de $MAE$ entre a estimativa hierárquica da idade com margens fixas e com regiões sobrepostas flexíveis em diferentes conjuntos de características (MORPH). ....	95
Figura 66	Idades utilizadas para as regiões sobrepostas flexíveis entre os SVRs (MORPH). ....	96
Figura 67	Desempenho em termos de $MAE$ dos regressores SVR da estimativa hie-	



	rárquica com regiões sobrepostas flexíveis em diferentes conjuntos de características (MORPH). . . . .	97
Figura 68	Escore acumulativo ( $CS$ ) dos métodos de estimativa da idade (MORPH). . . . .	98
Figura 69	Distribuição de idade após a combinação dos conjuntos de dados FG-NET e MORPH. . . . .	99
Figura 70	Desempenho da classificação em diferentes conjuntos de características (FG-NET+MORPH). . . . .	100
Figura 71	Matriz de confusão da classificação em faixas etárias utilizando a combinação de características AAM e $LPQ_{7 \times 7}$ (FG-NET+MORPH). . . . .	101
Figura 72	Comparação em termos de $MAE$ entre a estimativa hierárquica da idade com margens fixas e com regiões sobrepostas flexíveis em diferentes conjuntos de características (FG-NET+MORPH). . . . .	102
Figura 73	Idades utilizadas para as regiões sobrepostas flexíveis entre os SVRs (FG-NET+MORPH). . . . .	103
Figura 74	Desempenho em termos de $MAE$ dos regressores SVR da estimativa hierárquica com regiões sobrepostas flexíveis em diferentes conjuntos de características (FG-NET+MORPH). . . . .	104
Figura 75	Escore acumulativo ( $CS$ ) da estimativa hierárquica com regiões sobrepostas flexíveis (FG-NET+MORPH). . . . .	105
Figura 76	Comparação dos resultados da estimativa hierárquica da idade com regiões sobrepostas e a estimativa realizada por seres humanos em algumas imagens do conjunto FG-NET. . . . .	107
Figura 77	Histogramas dos códigos $LBP_{8,2}^{u2}$ para a região do canto direito da boca para 2, 17, 28 e 69 anos respectivamente (FG-NET). . . . .	119
Figura 78	Histogramas dos códigos $LBP_{8,2}^{u2}$ para a região das linhas da bochecha direita para 2, 17, 28 e 69 anos respectivamente (FG-NET). . . . .	120
Figura 79	Histogramas dos códigos $LBP_{8,2}^{u2}$ para a região das linhas do queixo para 2, 17, 28 e 69 anos respectivamente (FG-NET). . . . .	121
Figura 80	Histogramas dos códigos $LBP_{8,2}^{u2}$ para a região do topo do nariz para 2, 17, 28 e 69 anos respectivamente (FG-NET). . . . .	122
Figura 81	Histogramas dos códigos $LBP_{8,2}^{u2}$ para a região das linhas nasolabial direitas para 2, 17, 28 e 69 anos respectivamente (FG-NET). . . . .	123
Figura 82	Histogramas dos códigos $LPQ_{5 \times 5}$ para a região do canto direito da boca e suas respectivas orientações para 2, 17, 28 e 69 anos respectivamente (FG-NET). . . . .	125
Figura 83	Histogramas dos códigos $LPQ_{5 \times 5}$ para a região das linhas da bochecha direita e suas respectivas orientações para 2, 17, 28 e 69 anos respectivamente (FG-NET). . . . .	126
Figura 84	Histogramas dos códigos $LPQ_{5 \times 5}$ para a região das linhas do queixo e suas respectivas orientações para 2, 17, 28 e 69 anos respectivamente (FG-NET). . . . .	127
Figura 85	Histogramas dos códigos $LPQ_{5 \times 5}$ para a região do topo do nariz e suas respectivas orientações para 2, 17, 28 e 69 anos respectivamente (FG-	

	NET). . . . .	128
Figura 86	Histogramas dos códigos $LPQ_{5 \times 5}$ para a região das linhas nasolabial direita e suas respectivas orientações para 2, 17, 28 e 69 anos respectivamente (FG-NET). . . . .	129

## Lista de Tabelas

Tabela 2	Síntese e comparação de técnicas de estimação etária. Utilizada a tabela de Fu et al. [2] como referência. ....	21
Tabela 3	Funções kernel básicas .....	47
Tabela 4	Pontos de referência facial ( <i>landmarks</i> ) de acordo com a Figura 27. ...	53
Tabela 5	Distribuição dos dados de treinamento e teste para a estimativa <i>single-level</i> (FG-NET). ....	81
Tabela 6	Cinco melhores resultados para o método <i>Single-level</i> (FG-NET). ....	81
Tabela 7	Faixas etárias e distribuição dos dados de treinamento e teste para cada classe (FG-NET). ....	82
Tabela 8	Resultados de diferentes métodos de estimação etária em diferentes conjuntos de características, mostrando também a dimensão original das características e depois da redução da dimensionalidade utilizando PCA (FG-NET). ....	84
Tabela 9	<i>MAE</i> médio entre todos os conjuntos de características para algumas regiões sobrepostas fixas (FG-NET). ....	86
Tabela 10	Comparação em termos de dimensão e <i>MAE</i> (idade) utilizando a concatenação das características e a aplicação do método de fusão de características (FG-NET). ....	90
Tabela 11	Comparação dos resultados da estimativa hierárquica com margens flexíveis utilizando o protocolo dos experimentos anteriores (75% para treinamento e 25% para teste) e o LOPO (FG-NET). ....	91
Tabela 12	Distribuição do conjunto de dados em décadas (MORPH). ....	92
Tabela 13	Faixas etárias e distribuição dos dados de treinamento e teste para cada classe (MORPH). ....	92
Tabela 14	Resultados de diferentes métodos de estimação etária em diferentes conjuntos de características (MORPH). ....	96
Tabela 15	Distribuição dos dados de treinamento e teste para cada regressor SVR da estimativa específica da idade (FG-NET+MORPH). ....	100
Tabela 16	Resultados da estimativa hierárquica com margens fixas e flexíveis em diferentes conjuntos de características, mostrando também a dimensão das características após a redução da dimensionalidade utilizando PCA (FG-NET+MORPH). ....	102
Tabela 17	Comparação entre o desempenho dos seres humanos e os regressores SVR em estimar a idade aparente em determinadas faixas etárias utilizando o conjunto FG-NET .....	106
Tabela 18	Comparação entre o desempenho dos seres humanos em estimar a idade aparente e o método proposto de estimativa hierárquica da idade com regiões sobrepostas flexíveis utilizando o conjunto FG-NET. ....	106
Tabela 19	Comparação do método proposto com outros trabalhos reportados no	

	conjunto de dados FG-NET. ....	108
Tabela 20	Comparação do método proposto com outros trabalhos reportados no conjunto de dados MORPH. ....	109
Tabela 21	Distribuição dos dados de treinamento e teste para cada faixa etária do estimador hierárquico por décadas (FG-NET). ....	130
Tabela 22	Resultados da estimativa hierárquica por décadas em diferentes conjuntos de características (FG-NET). ....	132
Tabela 23	Resultados de diferentes métodos de estimação etária em diferentes conjuntos de características, mostrando também a dimensão original das características e depois da redução da dimensionalidade utilizando PCA (FG-NET). ....	134
Tabela 24	Resultados de diferentes métodos de estimação etária em diferentes conjuntos de características, mostrando também a dimensão após a redução da dimensionalidade utilizando PCA (MORPH). ....	135

## Lista de Siglas

AAM	Active Appearance Model
AG	Algoritmos Genéticos
AGES	AGing pattErn Subspace
ALR	Age-group-specific Linear Regression
ASM	Active Shape Model
BIF	Biologically Inspired Features
CEA	Conformal Embedding Analysis
CS	Cumulative Score
DFT	Discrete Fourier Transform
EM	Expectation-Maximization
FFT	Fast Fourier Transform
GMM	Gaussian Mixture Model
GW	Gabor Wavelets
KNN	k-Nearest Neighbors
LARR	Locally Adjusted Robust Regression
LBP	Local Binary Pattern
LOPO	Leave-One-Person-Out
LPP	Locality Preserving Projections
LPQ	Local Phase Quantization
MAE	Mean Absolute Error
OLPP	Orthogonal Locality Preserving Projections
PCA	Principal Component Analysis
PSF	Point Spread Functions
RBF	Radial Basis Function
RMIR	Robust Multi-instance Regression
RNN	Redes Neurais Artificiais
SFP	Spatially Flexible Patch
SOM	Self-Organizing Map
STFT	Short-Term Fourier Transform
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SVR	Support Vector Regression
WLD	Weber's Law Descriptor
YGA	Yamaha Gender and Age

## Resumo

A face humana contém uma grande quantidade de informações relacionadas com as características pessoais, incluindo identidade, gênero, etnia, estado emocional, idade etc. Atualmente, a idade desempenha um papel importante em muitas aplicações, tais como segurança, vigilância, interação homem-computador entre outras. O envelhecimento é um processo muito complexo, uma vez que as pessoas envelhecem de formas distintas que depende de diversos fatores (i.e., genética, estilo de vida etc.) que tornam o processo de determinar a idade de uma pessoa através da face muito desafiador. Diversas abordagens para a estimativa da idade têm sido exploradas considerando diferentes aspectos do rosto humano, tais como relações geométricas entre os principais elementos da face, formas dos elementos faciais ou texturas da pele. Os métodos de estimativa da idade são geralmente classificados segundo duas abordagens principais: a representação do envelhecimento facial e o método de estimativa da idade. Os modelos mais comuns para a representação do envelhecimento são os antropométricos, de aparência ativa (AAM), subespaço do padrão de envelhecimento (AGES), *age manifold* e modelos de aparência. Os métodos de estimativa da idade podem ser abordados como problemas de classificação multiclasse ou problemas de regressão. Este trabalho apresenta uma nova abordagem que modela o processo de envelhecimento através de características híbridas compostas por características globais (forma e aparência) e locais (rugas e manchas na pele). Várias técnicas de extração de características são analisadas, como modelos de aparência ativa, *Local Binary Pattern* (LBP), *Local Phase Quantization* (LPQ) e *Gabor wavelets*. Além disso, este trabalho propõe um novo método hierárquico para a estimação da idade composto por uma etapa de classificação em faixas etárias utilizando máquina de vetores de suporte e para a estimativa específica da idade utiliza-se máquina de vetores de suporte para regressão. O erro gerado na classificação de faixas etárias causado pelos limites rígidos do classificador é compensado na estimativa específica da idade pela sobreposição flexível de faixas etárias. Os experimentos para avaliar a abordagem proposta foram conduzidos em conjuntos de dados disponíveis publicamente FG-NET *Aging* e MORPH *Album 2*. A combinação de características globais e locais, AAM+LPQ<sub>7×7</sub>, foi a que obteve o melhor desempenho no conjunto FG-NET, tendo um erro médio absoluto de 4,50 anos. Para o conjunto MORPH o erro médio absoluto foi de 5,86 anos utilizando as características locais extraídas pelo operador LPQ<sub>7×7</sub>. A combinação dos conjuntos FG-NET e MORPH também foi avaliada para testar a robustez do método em um conjunto de dados maior e foi obtido um erro médio absoluto de 5,20 anos através da combinação AAM+LPQ<sub>7×7</sub>. O método proposto é comparado favoravelmente com a literatura existente e mostra ser robusto a variações de iluminação e borramentos devido o uso de características de fase local extraídas pelo operador LPQ.

Palavras-chave: Aprendizagem de máquina; Extração de características; Envelhecimento facial; Estimativa da idade; Características híbridas; Classificador hierárquico.

## Abstract

The human face conveys important perceptible information related to personal characteristics, including identity, gender, ethnicity, emotional state, age etc. Currently, the age plays an important role in many applications, such as security, surveillance, human computer interaction and others. Aging is a very complex process since people age in different ways which depends on many factors (i.e., genetics, lifestyle etc.) that make determining a person's age from the face very challenging. Several approaches for computer based age estimation have explored different aspects of human faces such as geometric ratios between major face features, shapes of facial parts or skin textures. The age estimation methods are classified according to the age image representation and the age prediction method. The most common models for the age image representation are Anthropometric models, Active Appearance Models (AAM), Aging Pattern Subspace (AGES), Age Manifold and Appearance Models. For age prediction, the problem is either tackled as a multi-class classification problem or a regression problem. This work presents a novel approach that models the process of aging using hybrid features composed by global (shape and appearance) and local (wrinkle and skin spots) features. Several features extraction approaches are analysed, such as Active Appearance Model (AAM), Local Binary Pattern (LBP), Local Phase Quantization (LPQ) and Gabor wavelets. In addition, this work proposes a novel hierarchical age estimation consisting of age group classification using Support Vector Machine (SVM) and specific age estimation using Support Vector Regression (SVR). The error in the age group classification caused by the hard boundaries is compensated in the detailed age estimation by overlapping the age ranges of each age function. Experiments have been carried out on two publicly available FG-NET Aging and MORPH Album 2 datasets to evaluate the performance of the proposed approach. The global and local features fusion, AAM+LPQ<sub>7×7</sub>, achieved 4,50 years of mean absolute error on the FG-NET dataset. The mean absolute error on the MORPH dataset was 5,86 years through the local features extracted by the LPQ<sub>7×7</sub> operator. In addition, the combination of FG-NET and MORPH datasets was evaluated in order to verify the robustness of the proposed approach and a mean absolute error of 5,20 years was achieved using the features combination AAM+LPQ<sub>7×7</sub>. The proposed approach compares favorably to the literature with increased robustness to lighting and blur variance through local phase features extracted by the LPQ operator.

Keywords: Machine learning; Features extraction; Facial aging; Age estimation; Hybrid features; Hierarchical classification.

# 1 Introdução

A partir de uma imagem facial humana é possível adquirir uma grande variedade de informações como identidade, idade, gênero, etnia, expressões emocionais, cicatrizes, marcas e tatuagens, como exemplificados na Figura 1. Tais informações têm sido largamente utilizadas no desenvolvimento de sistemas de interação humano-computador capazes de interpretar as informações pessoais encontradas nos humanos a partir de imagens e tomar decisões.

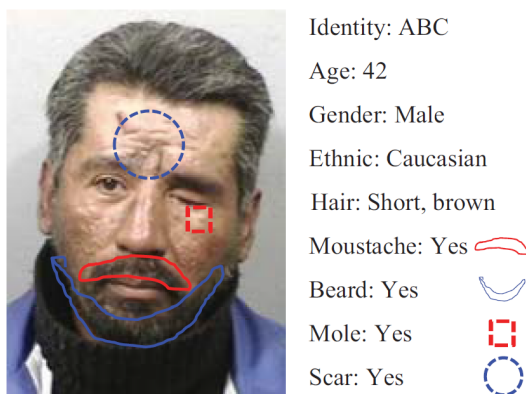


Figura 1: Exemplos de informações faciais [1].

Apesar da grande exploração da identificação de pessoas a partir de imagens faciais existe uma quantidade limitada de pesquisas em como estimar precisamente e usar informações contidas nas imagens faciais como idade, gênero e etnia [14]. Muitas aplicações práticas que precisam dos seres humanos para fornecer informações demográficas a partir de imagens faciais não são viáveis, portanto, o interesse na extração automática de características demográficas em imagens faciais vem crescendo muito. A classificação etária tem como objetivo atribuir um rótulo a um indivíduo a partir de uma imagem facial considerando a idade exata ou uma faixa etária, sendo um problema desafiador uma vez que a modelagem do envelhecimento humano é um processo complexo, pois a aparência de uma face pode variar por fatores naturais intrínsecos (e.g., genética) e extrínsecos (e.g., ambiente de trabalho, estilo de vida e saúde), além dos problemas de aquisição



das imagens como a posição, iluminação, expressão e de outros fatores como maquiagem, oclusões, degradação da imagem causada por ruído, plásticas etc.

O foco dessa pesquisa é a estimativa da idade de uma face a partir de imagem frontal e é dividida em duas etapas: representação do envelhecimento facial e estimativa da idade. A representação do envelhecimento facial consiste na extração de características faciais e é a etapa mais importante uma vez que irá afetar diretamente a etapa da estimativa da idade. As características podem ser categorizadas em locais, globais e também em características híbridas, sendo a combinação das locais e globais. As locais consistem na quantidade e profundidade das rugas, sardas e outras marcas, cor do cabelo e geometria facial. E têm sido utilizadas para classificar pessoas em grupos etários (e.g., bebês, jovens, adultos e idosos), pois possuem características de determinados grupos, como rugas por exemplo, encontradas normalmente em pessoas mais velhas. As características globais são geralmente utilizadas para fazer um detalhamento da idade e não contêm apenas características relacionadas ao envelhecimento, mas também características individuais como identidade, expressão, gênero, etnia etc.

A etapa da estimativa da idade pode ser dividida em três abordagens: classificação de faixas etárias [15], estimativa *single-level* [16] e classificadores hierárquicos [12]. A classificação de faixas etárias coloca o indivíduo num determinado grupo enquanto o *single-level* e o hierárquico são construídos para prever uma idade mais precisa. A estimativa automática da idade possui diversas aplicações: (i) Justiça: Podendo ajudar a determinar potenciais suspeitos com maior precisão e eficiência filtrando um banco de dados de imagens utilizando uma idade estimada como fator de busca. (ii) Controle de segurança: Proibindo menores de idade de comprar bebidas alcoólicas e cigarros em máquinas automáticas e também de acessar conteúdos impróprios na Internet. (iii) Interação humano-computador: Mostrando conteúdos de acordo com a faixa etária do usuário, por exemplo, uma televisão que ajusta a programação de acordo com a idade. (iv) Sistemas de reconhecimento facial robustos a progressão de idade.

## 1.1 Definição do problema

Apesar do relativo sucesso em alguns cenários controlados (imagens de alta-qualidade, sem oclusão, sem fundo e expressões faciais neutras) a maioria dos métodos atuais não apresentam um bom desempenho em ambientes livres de restrições, como por exemplo, câmeras de segurança de baixa resolução para aplicações de controle de acesso utilizando imagens

faciais. Portanto, nesta pesquisa são utilizadas as bases de dados públicas FG-NET *Aging* e MORPH *Album 2* que possuem variações de expressão, iluminação, resolução, posição e ruído, deixando o modelo mais próximo de aplicações reais.

A estimativa automática da idade compartilha similaridades com problemas de progressão da idade que prediz a aparência facial futura baseada em imagens com as aparências anteriores de um indivíduo. Devido ambas levarem em conta os efeitos do envelhecimento durante o ciclo da vida, pesquisas na área da estimativa da idade podem contribuir para os problemas de progressão e vice versa. Um importante aspecto no problema da estimação etária é a formulação de métricas adequadas para avaliar o desempenho dos sistemas. O erro médio absoluto (MAE, do inglês *Mean Absolute Error*) é a métrica mais utilizada, mostrando a média das diferenças absolutas entre idades estimadas e reais, também, o escore acumulativo (CS, do inglês *Cumulative Score*) que mostra o percentual de casos em um conjunto de dados de teste onde o erro é menor que um limiar.

A estimação etária é um problema desafiador, visto que diversos fatores podem influenciar no processo de envelhecimento facial como genética, etnia, gênero e estilo de vida. Além dos desafios relacionados a diferentes perturbações presentes nas imagens faciais como variações na iluminação, expressão facial, pose e borramentos nos quais não tem sido bem explorados em trabalhos anteriores.

## 1.2 Hipóteses de pesquisa

A hipótese a ser verificada neste trabalho é a de que seja possível representar o envelhecimento facial através da combinação de características globais e locais que sejam robustas a variações de iluminação, pose, expressão facial e borramentos. Assim como a hipótese de que a estimativa da idade utilizando um classificador hierárquico contendo uma etapa de classificação em faixas etárias, e outra para a estimativa específica da idade a partir de regressores com regiões sobrepostas flexíveis entre as classes se possa alcançar resultados similares ou superiores aos alcançados por métodos atuais.

## 1.3 Objetivos

### 1.3.1 Objetivo geral

A pesquisa tem como objetivo o desenvolvimento de um método para a estimação da idade a partir de imagens faciais onde serão avaliadas técnicas de extração de características faciais híbridas baseadas em modelos de aparência ativa (AAM) para a extração de características globais e *Gabor Wavelets* (GW), *Local Binary Pattern* (LBP) e *Local Phase Quantization* (LPQ) para a extração de características locais, assim como avaliar a técnica de aprendizagem máquinas de vetor de suporte (SVM, do inglês *Support Vector Machine*) para a classificação de um indivíduo em uma faixa etária e a técnica de regressão de vetor de suporte (SVR, do inglês *Support Vector Regression*) para uma estimativa mais precisa da idade assumindo um ambiente não controlado onde as imagens podem conter diferentes perturbações como variações na iluminação, expressão facial, pose e borramentos. A partir do modelo gerado a validação será realizada utilizando dados de benchmark e métricas estatísticas para verificar o desempenho e aproximação com outros sistemas propostos.

### 1.3.2 Objetivos específicos

- Definição do conjunto de dados com imagens faciais;
- Implementação e análise de extratores de características globais e locais;
- Implementação e análise de algoritmos de classificação e regressão;
- Desenvolvimento de uma abordagem hierárquica para a estimação etária;
- Teste e avaliação sobre um conjunto de dados de benchmark para permitir uma comparação com outros trabalhos.

## 1.4 Justificativa

Nos últimos anos muitos pesquisadores mostraram métodos de estimação etária nos quais os desempenhos são comparáveis ou superiores às habilidades humanas. Uma das principais razões da necessidade de desenvolver sistemas automáticos é que os seres humanos não são tão precisos ao estimar a idade de um determinado indivíduo e futuramente será necessário desenvolver sistemas que tenham um desempenho superior considerável.

A estimativa automática da idade é ainda um problema desafiador. Diversos fatores podem influenciar no processo de envelhecimento incluindo etnia, gênero, genética e estilo de vida. Idealmente, os estimadores automáticos da idade devem operar sem restrições em imagens faciais para poderem ser utilizados em aplicações reais não importando variações de expressão, iluminação, resolução, posição, oclusões e borramentos.

Apesar do relativo sucesso em alguns cenários controlados (imagens de alta-qualidade, sem oclusões, sem fundo e expressões faciais neutras) a maioria dos métodos atuais não apresentam um bom desempenho em ambientes livres de restrições. Portanto, como realizar uma estimativa etária que seja robusta a todos esses fatores continua a ser um problema em aberto e este trabalho visa contribuir para aumentar a robustez da estimativa etária a estas variações apresentando resultados em termos de  $MAE$  e  $CS$  melhores ou comparáveis ao da literatura.

## 1.5 Contribuições

Com o desenvolvimento deste trabalho importantes contribuições no contexto da estimativa etária a partir de uma imagem facial foram obtidas. Entre elas, a apresentação de um novo método de estimativa hierárquica da idade com regiões sobrepostas flexíveis utilizando características extraídas a partir da técnica LPQ sendo robusto não apenas ao borramento mas também a outros fatores como variações na iluminação e expressão facial, muito presente nos conjuntos de dados utilizados e pouco explorados nos trabalhos anteriores. Adicionalmente, o método proposto foi avaliado no conjunto de dados MORPH *Album 2* o qual tem sido pouco explorado, assim como uma combinação dos conjuntos FG-NET e MORPH foi realizada para analisar a robustez do método através de um conjunto de dados maior. Os resultados obtidos foram suficientes para confirmar a hipótese de que a representação do envelhecimento facial através da combinação de características globais (AAM) e locais (LPQ) que sejam robustas a variações de iluminação, pose, expressão facial e borramentos tem um desempenho superior quando comparado às técnicas tradicionais utilizadas em outros trabalhos como LBP e Gabor *wavelets* [12]. Assim como a confirmação da hipótese de que a realização da estimativa etária utilizando uma hierarquia de classificadores especializado em faixa etária e estimativa da idade específica a partir de regressores com regiões sobrepostas flexíveis possui resultados similares ou superiores aos alcançados por métodos atuais.

## 1.6 Organização

Este trabalho está dividido em seis seções principais seguidos pelas referências e apêndices. A primeira seção, *Introdução*, explica a motivação por trás do trabalho, as hipóteses de pesquisa, diferenças entre pesquisas similares já realizadas, as principais contribuições e quais são os objetivos em vista. A seção *Revisão bibliográfica* mostra o estado da arte sobre a estimativa da idade e seus métodos. Na seção *Fundamentação teórica* são apresentados os principais fundamentos teóricos inerentes as técnicas utilizadas para a realização dos experimentos. A seção *Método proposto* descreve o método proposto para a construção do sistema de estimativa hierárquica da idade utilizando características globais e locais. Na seção *Resultados experimentais* são apresentados os resultados obtidos utilizando o método proposto, resultados de outros métodos de estimação para efeitos de comparação, assim como comparações com resultados de outros trabalhos. A última seção, *Conclusão*, fecha o trabalho com um resumo geral do que foi desenvolvido, as principais contribuições e sugestões para trabalhos futuros.

## 2 Revisão bibliográfica

Um grande número de pesquisas nas áreas biológicas, psicológicas e ciências cognitivas têm sido realizados para tentar entender como o cérebro humano percebe, representa e lembra as faces. Principalmente, a estimativa da idade feita por humanos estudados no campo da psicologia para examinar os efeitos da idade, gênero e etnia [17], as quais contribuíram muito para os métodos automáticos mostrados na literatura científica.

A aparência da face humana é afetada consideravelmente pelo processo de envelhecimento devido principalmente ao crescimento e movimento ósseo e deformações na pele causado pelo enfraquecimento muscular e rugas, como exemplificados na Figura 2 que mostra o envelhecimento facial de Albert Einstein da idade adulta até a sua velhice. As observações das características relacionadas ao envelhecimento permitem que os humanos estimem a idade de outras pessoas apenas olhando em suas faces. Porém, pesquisas relacionadas ao processo de estimação humana concluíram que humanos não são tão precisos em estimar a idade de outra pessoa mostrando um caminho atrativo para o desenvolvimento de técnicas automáticas [17].

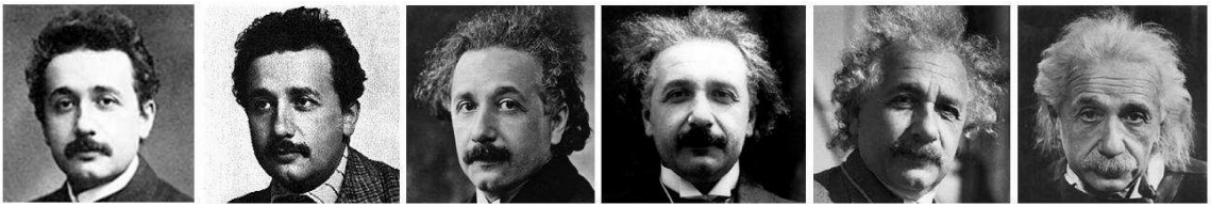


Figura 2: Exemplo da faixa etária de Albert Einstein [2].

Os métodos atuais para a classificação etária a partir de imagens faciais diferem em dois aspectos: na representação do envelhecimento facial e na abordagem de algoritmos utilizados para a estimaco etária. Diversos métodos de representao de imagens faciais têm sido explorados como modelos antropométricos [4], modelos de aparncia ativa (*Active Appearance Model* - AAM) [16], subespaço do padro de envelhecimento (*AGing pattErn Subspace* - AGES) [5], *age manifold* [6] e modelos de aparncia [12]. Com relao às

abordagens de estimaco etria os mtodos existentes so basicamente os classificadores convencionais como redes neurais [18], mquina de vetores de suporte (*Support Vector Machine* - SVM) [5] ou ento anlise de regresso.

A coleta de imagens faciais para a estimaco etria  crucial para o desenvolvimento de tais mtodos, especialmente imagens de um mesmo indivduo em uma srie cronolgica. As bases de dados que mais tm sido utilizadas para o treinamento e avaliao da estimativa da idade so: FG-NET *Aging*, MORPH *Album 2* e YGA, conforme mostradas na Tabela 2 no final da seo. A FG-NET *Aging*  uma base de dados pblica contendo 1.002 imagens faciais coloridas e em escala de cinza de 82 pessoas de 0 a 69 anos de mltiplas etnias, e com grande variao de iluminao, pose e expresso facial. A base de dados MORPH *Album 2*, tambm disponibilizada publicamente, contm mais de 20.000 imagens faciais obtidas a partir de mais de 4.000 indivduos na faixa etria de 16 a 77 anos, principalmente de afro-descendentes e caucasianos. A YGA (*Yamaha Gender and Age*)  uma base de dados privada que contm 8.000 imagens coloridas de 1.600 indivduos asiticos, sendo 800 homens e 800 mulheres variando de 0 a 93 anos em diversas condioes de iluminao, pose e expresso facial.

## 2.1 Representao do envelhecimento facial

### 2.1.1 Modelos antropomtricos

A antropometria facial  a cincia para medir os tamanhos e proporoes da face humana. Uma das primeiras pesquisas sobre estimaco etria foi reportada por Kwon e Lobo [4] onde utilizaram um modelo antropomtrico baseado na distncia dos elementos faciais para extrair caractersticas geomtricas e rugas para classificar imagens faciais entre bebs, adultos ou idosos. A principal ideia dessa abordagem  utilizar pesquisas sobre o desenvolvimento crnio-facial que utilizam modelos matemticos. O modelo de transformao de tenso cardioidal revisado descreve o crescimento do crnio a partir da analogia entre o crescimento da cabea humana e as pressoes exercidas em um objeto esfrico preenchido com fluido, podendo assim fazer uma anlise hidrosttica dos efeitos da gravidade no crescimento da cabea. Matematicamente o modelo cardioidal pode ser expresso como  $P \propto R_0(1 - \cos(\theta_0))$ ,  $R_1 = R_0 + k(R_0 - R_0 \cos \theta_0)$ ,  $\theta_1 = \theta_0$ , onde  $P$  corresponde a presso aplicada na face,  $(R_0, \theta_0)$ ,  $(R_1, \theta_1)$  correspondem as coordenadas polares da face antes e aps a transformao e  $k$   o parmetro que aumenta ao longo do tempo. A Figura 3 ilustra os efeitos do modelo de transformao cardioidal na induo de alteraoes

de crescimento da cabeça de uma criança. Farkas [19] forneceu uma visão abrangente da antropometria facial definindo-a em termos de medidas tiradas a partir de 57 pontos de referência em faces humanas de diferentes idades desde a infância até a idade adulta. Para uma representação do crescimento de acordo com a idade geralmente são utilizadas distâncias e proporções medidas a partir de pontos de referência ao invés de utilizar o modelo matemático descrito anteriormente. Existem duas razões que as pessoas não utilizam tais modelos matemáticos para uma estimativa etária: 1) O modelo matemático não pode representar o perfil da cabeça de uma forma natural, especialmente na idade adulta e 2) o perfil da cabeça é difícil de medir a partir de imagens faciais bidimensionais.

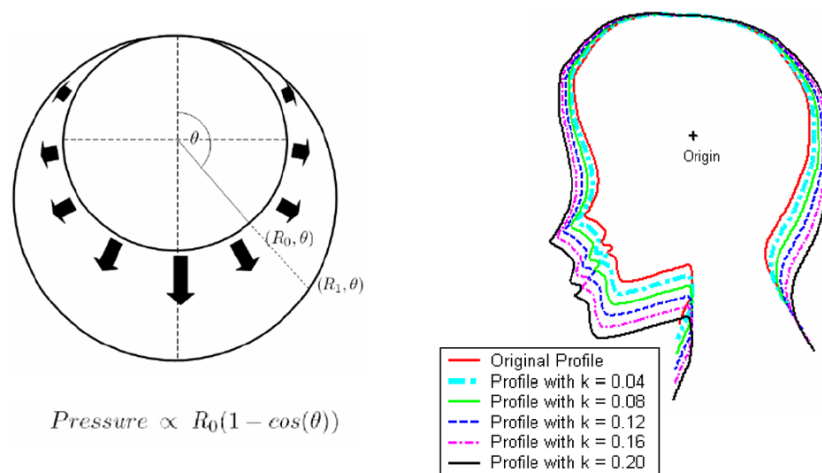


Figura 3: Distribuição da pressão dentro de um objeto esférico preenchido com fluido e crescimento facial simulado no perfil da cabeça de uma criança utilizando o modelo de transformação de tensão cardioidal revisado [3].

Na representação da idade a partir de imagens faciais Kwon e Lobo [4] computaram seis proporções de distância em imagens frontais da face conforme demonstrado na Figura 4 para diferenciar bebês de adultos e utilizaram padrões de rugas para diferenciar adultos de idosos. As rugas foram detectadas em regiões próximas aos olhos, testa e bochechas, conforme Figura 5(a), e extraídas utilizando *snakelets* [4] que detectam a presença de curvas como as rugas nas regiões determinadas, conforme mostrado na Figura 5(b). O experimento foi conduzido em uma pequena base de dados com 47 imagens faciais.

Ramanathan e Chellappa [3] utilizaram oito proporções de distância para modelar a progressão de idade em faces jovens (e.g., 0 a 18 anos). O objetivo é prever a aparência de um indivíduo através da idade e realizar o reconhecimento facial através da progressão da idade. Seus experimentos foram conduzidos em um conjunto de dados contendo 233 imagens de 109 indivíduos parcialmente coletados da base de dados FG-NET *Aging* e de



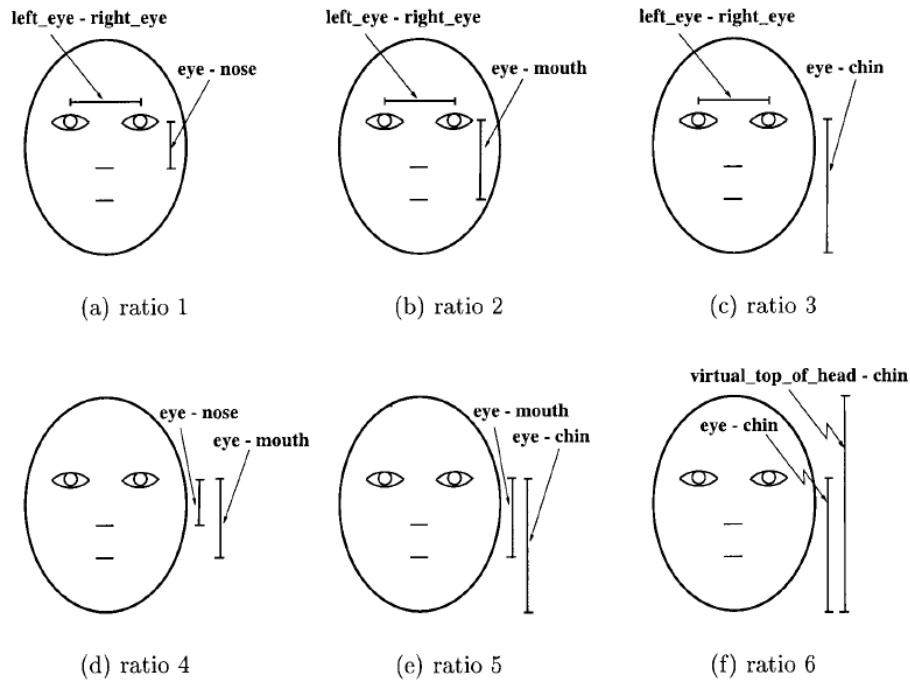


Figura 4: As seis proporções de distância em imagens frontais da face [4].

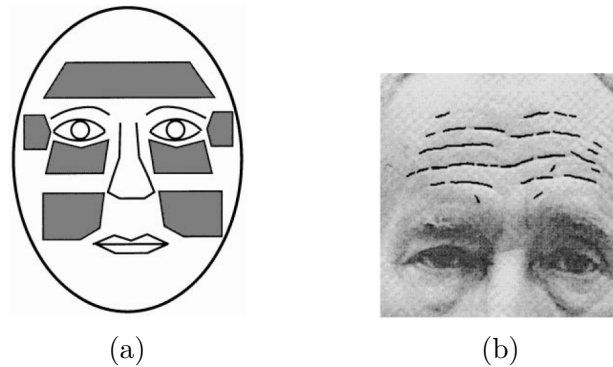


Figura 5: (a) Regiões da face contendo rugas faciais e (b) definição das rugas na região da testa encontradas utilizando *snakelets* [4].

base de dados própria, obtendo uma melhoria para o reconhecimento facial ao utilizar a progressão da idade de 8% (sem prever idade) para 15% (considerando a progressão da idade).

Abordagens de estimativa da idade baseadas em modelos antropométricos podem apenas tratar idades jovens uma vez que o formato da cabeça não muda muito em idades adultas. Em prática, somente faces frontais podem ser utilizadas para a medição das geometrias faciais devido as proporções de distância serem computadas a partir de imagens 2D nas quais as proporções são sensíveis ao posicionamento da face. Finalmente, a representação utilizando modelo antropométrico considera apenas a geometria facial e nenhuma informação de textura, além do modelo proposto por Farkas ser baseado nas

medidas faciais de europeus não considerando indivíduos de outras etnias.

### 2.1.2 Modelos de aparência ativa

O modelo de aparência ativa (AAM) é um modelo estatístico da face proposto inicialmente por Cootes [9] para a representação de imagens faciais. Dado um conjunto de imagens faciais para treinamento, um modelo estatístico do formato da face e um modelo de intensidades são aprendidos separadamente baseados na análise dos componentes principais (PCA, do inglês *Principal Component Analysis*). Lanitis et al. [20], estenderam os AAMs para uma representação do envelhecimento propondo uma função,  $age = f(b)$ , para explicar as variações na idade. Na função,  $age$  é a idade atual de um indivíduo em uma imagem facial,  $b$  é um vetor contendo 50 parâmetros aprendidos a partir do AAM e  $f$  é a função do envelhecimento facial onde define a relação entre a idade e os parâmetros do modelo.

Os experimentos foram realizados em um conjunto de dados de 500 imagens faciais de 60 indivíduos onde 45 possuem imagens relacionadas à progressão da idade com variações de 1 a 30 anos de idade. Eles mostraram que o modelo pode melhorar o desempenho de reconhecimento facial de 67% para 71%. Em comparação com os modelos antropométricos os AAMs podem tratar qualquer idade em geral ao invés de jovens apenas. Também, AAMs consideram o formato facial e a textura ao contrário dos modelos antropométricos que consideram apenas a geometria facial.

Os modelos de aparência ativa são principalmente utilizados para estimar a idade a partir de características globais, pois oferecem melhores informações relativas à aparência e formato da face humana quando comparados às características locais. Porém, não incluem informações detalhadas de rugas e pele devido à redução da dimensão do espaço de características feita pela análise de componentes principais [12].

### 2.1.3 Subespaço do padrão de envelhecimento

Ao invés de tratar cada face separadamente para a estimativa etária uma sequência de imagens faciais de um indivíduo em diferentes idades podem ser usadas juntas para criar um modelo do processo de envelhecimento. Essa ideia foi explorada por Geng et al. [5] e é chamada de *AGing pattErn Subspace* (AGES). Um padrão do envelhecimento é definido como uma sequência de imagens faciais de um mesmo indivíduo em uma ordem temporal. O exemplo do padrão de envelhecimento (*aging pattern*) ilustrado na Figura 6 mostra

cada idade do indivíduo (0 a 8 anos) sendo alocada em uma posição ao longo do eixo  $t$ . Caso imagens faciais estão disponíveis para certas idades (2, 5 e 8 anos), elas são colocadas nas determinadas posições. Caso contrário, as posições são deixadas em branco. Se todas as posições são preenchidas o padrão é chamado de *padrão de envelhecimento completo*, senão, é chamado de *padrão de envelhecimento incompleto*.

Antes do padrão de envelhecimento ser processado, as imagens faciais são transformadas em vetores de características. Como o processo de envelhecimento é relacionado a mudanças no formato e textura facial o modelo de aparência ativa é utilizado para a extração das características. A Figura 6 mostra um exemplo da vetorização do padrão de envelhecimento (*aging pattern vector*), onde  $b_2$ ,  $b_5$  e  $b_8$  representam os vetores de características das faces nas idades 2, 5 e 8 anos, respectivamente. Um modelo representativo para os padrões de envelhecimento pode ser construído a partir da técnica PCA para capturar as principais variações nos conjuntos de dados. A dificuldade é que o padrão de envelhecimento é normalmente incompleto, pois é muito difícil obter imagens faciais de um mesmo indivíduo em uma ordem temporal completa. No entanto, o método AGES é capaz de sintetizar as imagens faciais nas idades faltantes através de uma aprendizagem iterativa [5] utilizada para minimizar o erro de reconstrução caracterizado pela diferença entre as imagens disponíveis e as imagens sintetizadas.

Os valores iniciais para a sintetização das imagens faltantes são definidos pela média das imagens faciais disponíveis. Então, os *eigenvectors* da matriz de covariâncias de todas as imagens e as médias podem ser computadas. Para a estimação da idade a imagem facial de teste precisa encontrar um padrão de envelhecimento apropriado para ele e também uma posição da idade adequada no padrão. A posição da idade é retornado como a idade estimada da imagem facial de teste. Para isso, a imagem de teste é verificada em cada posição possível no padrão de envelhecimento e aquela posição com o menor erro de reconstrução é selecionada.

O experimento para avaliar o desempenho do método foi realizado utilizando a base de dados FG-NET *Aging*. O erro médio absoluto reportado foi de 6,77 anos, sendo competitivo com outros trabalhos reportados na literatura, como pode ser comparado na Tabela 2 no final da seção. O problema no uso prático do método AGES é que para estimar uma idade de uma imagem teste o método assume que existem imagens do mesmo indivíduo em diferentes idades, ou pelo menos um padrão de envelhecimento similar para a imagem de teste que foi utilizada para o treinamento. Um outro problema é que quando a imagem é representada pelo AAM podem não serem representadas muito bem as rugas

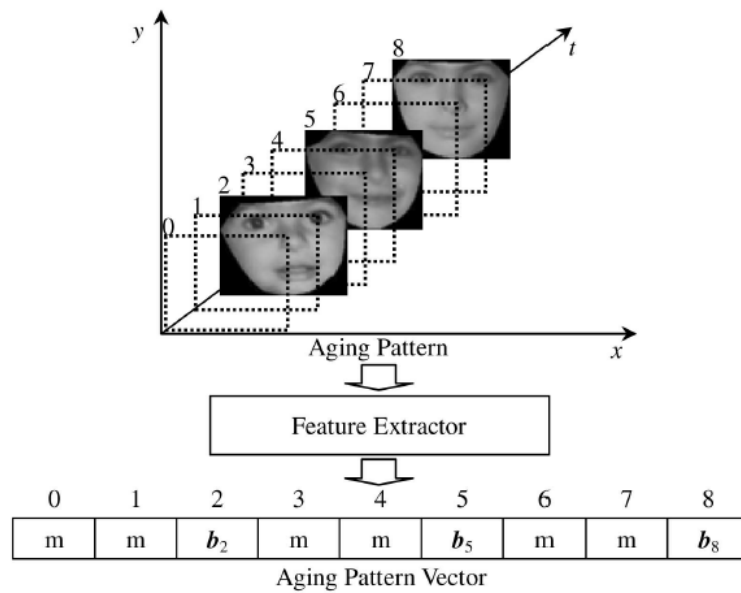


Figura 6: Vetorização do padrão de envelhecimento. As idades (0-8) são marcadas numa ordem temporal e o vetor de características é identificado com as imagens que estão disponíveis ( $b_i$ ) e a imagens faltantes são identificados como  $m$  [5].

de pessoas idosas, pois a imagem é codificada em termos de intensidades que geralmente não caracterizam informações locais de texturas.

#### 2.1.4 *Age manifold*

Ao contrário de aprender um padrão de envelhecimento específico para cada indivíduo, um padrão comum pode ser aprendido a partir de vários indivíduos em diferentes idades. Para cada idade várias imagens faciais podem ser utilizadas para representar tal idade. Cada indivíduo, homem ou mulher, pode ter imagens em apenas uma idade ou em uma faixa de idades. Esse método de aprendizagem de um padrão de envelhecimento faz com que a representação da idade seja muito flexível quando comparada ao método AGES. Uma maneira possível de aprender padrões comuns de envelhecimento é o método *age manifold* [6] o qual utiliza a técnica *manifold embedding* para aprender um padrão de envelhecimento a partir de várias imagens faciais em diversas idades e representá-las através de um espaço de características de baixa dimensão mantendo a geometria intrínseca do conjunto de dados. Não há condições para possuir várias faces em diferentes idades para cada indivíduo. Em casos extremos um indivíduo pode possuir imagens em apenas uma idade na base de dados. Essa flexibilidade faz com que a coleta da base de dados seja mais fácil.

Supondo que o espaço da imagem é amostrado por um conjunto de imagens faciais  $\mathcal{X} = \{x_i : x_i \in \mathbb{R}^D\}_{i=1}^n$  de  $m$  indivíduos com a dimensão dos dados  $D$  em ordem de idade. Um conjunto de *ground truth*  $\mathcal{L} = \{l_i : l_i \in \mathbb{N}\}_{i=1}^n$  associados com a imagem providenciam os labels da idade. O objetivo da aprendizagem *age manifold* é encontrar uma representação de baixa dimensão no subespaço capturando a distribuição intrínseca dos dados e uma estrutura geométrica, assim como sua representação  $\mathcal{Y} = \{y_i : y_i \in \mathbb{R}^d\}_{i=1}^n$  com  $d \ll D$ . Assim, o mapeamento a partir do espaço da imagem para o espaço *manifold* pode ser formulado utilizando uma função linear ou não-linear, tal como  $\mathcal{Y} = P(\mathcal{X}, \mathcal{L})$ . O *manifold embedding* pode ser calculado pela aprendizagem de uma matriz  $P$  de dimensão  $D \times d$  satisfazendo  $Y = P^T X$  de uma maneira linear ou aprendendo diretamente  $Y$  de uma maneira não-linear, onde  $Y = [y_1 \ y_2 \ \dots \ y_n]$ ,  $X = [x_1 \ x_2 \ \dots \ x_n]$  e  $P = [p_1 \ p_2 \ \dots \ p_d]$ . Um algoritmo eficiente, chamado projeção de localidade ortogonal preservada (OLPP, do inglês *Orthogonal Locality Preserving Projections*), serve para conectar a aprendizagem *manifold* com a aprendizagem do subespaço. O OLPP basicamente produz uma função de base ortogonal baseado no algoritmo projeção de localidade preservada (LPP, do inglês *Locality Preserving Projections*) onde busca uma incorporação para preservar as estruturas *manifolds* essenciais medindo as distâncias de seus vizinhos locais.

A Figura 7 ilustra a visualização dos resultados do *age manifold* que revelam que o método pode ser aprendido utilizando uma base de dados grande e demonstra que o OLPP pode representar bem o *age manifold*, enquanto o tradicional PCA não é muito eficiente para esse propósito. Fu e Huang [7] também mostraram que o método análise de incorporação conformável (CEA, do inglês *Conformal Embedding Analysis*) pode ser usado para melhorar o desempenho da estimativa da idade. Guo et al. [21] providenciaram mais evidências para mostrar que o *age manifold* aprendido com o método OLPP pode representar bem a idade de uma imagem e o *MAE* reportado foi de aproximadamente 5,30 anos em um grande conjunto de dados quando o método regressão robusta localmente ajustada (LARR, do inglês *Locally Adjusted Robust Regression*) foi utilizado. E o *MAE* no conjunto de dados FG-NET foi de 5,07 anos.

*Age manifold* é um método que necessita de uma análise do subespaço para encontrar padrões relacionados a idade. O método encontra um espaço embutido de baixa dimensão quando cada idade possui várias imagens faciais. Uma propriedade interessante na utilização dos métodos OLPP ou CEA para a aprendizagem é que os labels das idades podem ser incorporados no processo de uma maneira supervisionada. Enquanto o método PCA pode trabalhar apenas de uma maneira não-supervisionada sem informações da idade para aprender a representação do subespaço. *Age manifold* não necessita

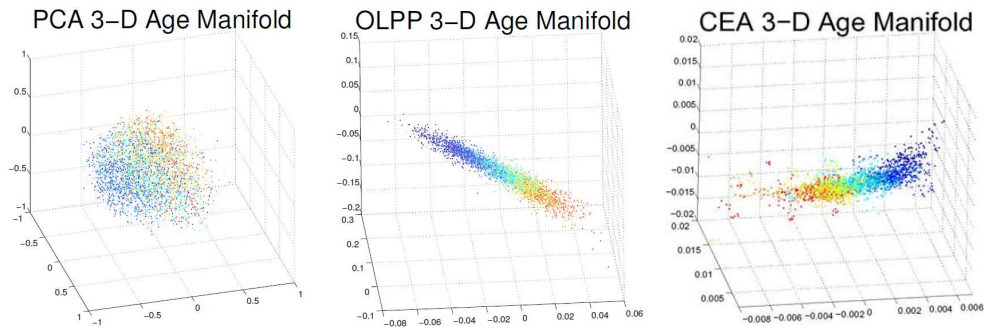


Figura 7: *Ageing manifold* utilizando PCA, OLPP e CEA [6] [7].

aprender um padrão de envelhecimento individual para cada indivíduo. Todas as idades de diferentes indivíduos podem ser usadas juntas, sendo que a única restrição é que o tamanho do conjunto de dados de treinamento seja grande o suficiente para aprender as diversidades das características faciais para representar o envelhecimento em um espaço de baixa dimensão.

### 2.1.5 Modelos de aparência

Características faciais relacionadas ao envelhecimento são mais realçadas pelos modelos de aparência. Características globais e locais têm sido utilizadas em sistemas de estimação etária. Hayashi et al. [22] consideraram textura (rugos) e formato (geometria) como características para representar cada imagem facial. O sistema estima a idade através de uma classificação multi-grupo com 5 anos de intervalo. Uma pequena base de dados japonesa com 300 indivíduos foi utilizada cobrindo idades de 15 a 64 anos. Estimação de gênero também foi implementada para melhorar o desempenho uma vez que os padrões de envelhecimento são diferentes para homens e mulheres. Fukai e Takimoto [23] utilizaram a transformada rápida de Fourier (FFT, do inglês *Fast Fourier Transform*) para extrair o espectro de características de uma aparência facial e utilizaram algoritmos genéticos (AG) para a seleção de características.

Os descritores de textura LBP e Gabor *wavelets* também têm sido utilizados para a extração de características relacionadas a aparência da face humana [12]. Yan et al. [24], [25] propuseram usar o descritor de características *Spatially Flexible Patch* (SFP) uma vez que considera manchas locais e suas posições. Imagens faciais com pequenos desalinhamentos, oclusões e variações de pose podem ser utilizadas efetivamente. Modelado por *Gaussian Mixture Model* (GMM) o sistema pode atingir um *MAE* de 4,95 anos na base de dados FG-NET. Guo et al. [26], propuseram utilizar características biologicamente inspi-

radas (BIF, do inglês *Biologically Inspired Features*) para a estimação etária via faces. A ideia básica é derivada de um modelo *feedforward* do sistema de reconhecimento visual de objetos dos primatas, o modelo "HMAX". O modelo possui camadas alternadas chamadas de células unitárias *Simples* (S) e *Complexas* (C) onde a complexidade aumenta conforme a progressão das camadas a partir do córtex visual primário (V1) para o córtex inferior temporal (IT). A primeira camada S1 é criada utilizando filtros de Gabor na imagem. A segunda camada C1 é obtida com uma operação "MAX" na S1. Essas características tem a vantagem de lidar com pequenas mudanças em translações, rotações e escalas. As características BIF combinadas com SVM podem atingir um *MAE* de 4,77 anos na base de dados FG-NET e 3,47 na base de dados YGA.

## 2.2 Algoritmos para estimação etária

Dada uma representação da idade facial o próximo passo é estimar a idade. A estimação etária pode ser vista com o uma tarefa de reconhecimento de padrões, mas de uma forma especial. Cada label de idade pode ser vista como uma classe, portanto, a estimação etária pode ser vista como um problema de classificação. De outro lado, a idade é uma sequência de valores numéricos, e.g., 0, 1, 2,...; portanto, a estimação etária pode ser vista também como um problema de regressão.

### 2.2.1 Classificação

Lanitis et al. [16] avaliaram o desempenho de diferentes classificadores para a estimação etária incluindo o classificador *k-Nearest Neighbors* (KNN), redes neurais artificiais (RNN) e um classificador de função quadrática. As características faciais foram extraídas pelo método AAM. A partir de experimentos em uma pequena base de dados contendo 400 imagens de 0 a 35 anos foi reportado que o classificador de função quadrática pode atingir 5,04 anos de *MAE* sendo um pouco menor que o classificador KNN, porém maior do que o classificador RNA e do que uma abordagem não-supervisionada, com o método *Self-Organizing Map* (SOM).

O SVM foi aplicado para a estimação etária por Guo et al. [21] na base de dados YGA com 8.000 imagens utilizando AAM para a representação facial. O *MAE* atingido foi de 5,55 e 5,52 anos para mulheres e homens, respectivamente. O *MAE* foi de 7,16 anos para a base de dados FG-NET. Kanno et al. [27] utilizaram RNA para uma classificação em quatro grupos etários o qual atingiu 80% de taxa de acerto em 110 faces

de homens jovens utilizando modelo de aparência para a representação facial através da técnica *mosaic* para a extração de características relacionadas a cor, tamanho e localização espacial dos elementos faciais.

### 2.2.2 Regressão

Considerando a estimação etária como um problema de regressão, Lanitis et al. [20] investigaram três formulações para a definição de uma função da idade: linear, quadrática e cúbica. Os parâmetros ótimos foram definidos a partir da aprendizagem do conjunto de treinamento em diferentes idades baseado nos algoritmos genéticos. O método SVR foi aplicado por Guo et al. [21] no *age manifold* OLPP para a estimação etária. O *MAE* atingido foi de 7 e 7,47 anos na bases de dados YGA para mulheres e homens, respectivamente, e 5,16 para a FG-NET.

O algoritmo *Expectation-Maximization* (EM) foi utilizado para resolver o problema de regressão e agilizar o processo de otimização [24]. As características foram extraídas a partir da técnica AAM e o *MAE* reportado foi de 6,95 anos para homens e mulheres da base YGA e 5,33 anos para a base FG-NET. Suo et al. [28] compararam *Age-group-specific Linear Regression* (ALR), *multilayer perceptrons*, SVR e regressão logística na base de dados FG-NET e em suas próprias bases de dados. Melhores resultados foram atingidos utilizando *multilayer perceptrons*. Ni et al. [29] propuseram o *Robust Multi-instance Regression* (RMIR) que pode lidar com labels discrepantes (*outliers*) quando construído um estimador etário universal baseado em uma base de dados grande com dados coletados da Web. O sistema pode atingir resultados razoáveis para a estimação etária com um treinamento automático online, obtendo um *MAE* de 8,60 anos.

### 2.2.3 Abordagens híbridas

O problema da estimação etária pode ser visto tanto como um problema de classificação quanto um problema de regressão. Portanto, classificadores e regressores representativos devem ser comparados em bases de dados iguais para avaliar as diferenças de desempenho de ambos. Guo et al. [21] escolheram SVM como um classificador representativo e SVR para um regressor representativo para estimação etária e compararam seus desempenhos utilizando os mesmos dados de entrada. A partir do experimento o SVM teve um desempenho muito superior do que o SVR na base de dados YGA (5,55 versus 7,00 e 5,52 versus



7,47 para homens e mulheres, respectivamente) enquanto o SVM teve um desempenho pior do que SVR na base de dados FG-NET (7,16 versus 5,16 anos). Portanto, é possível notar que utilizando a classificação para a estimação etária pode ser muito pior ou muito melhor do que a regressão em diferentes casos.

Uma maneira de melhorar a robustez é combinar os métodos de classificação e regressão para tirar vantagens dadas por ambos. Baseado nessa motivação, Guo et al. [21] propuseram o método chamado LARR. Eles mostraram que um desempenho considerável pode ser obtido a partir da combinação classificação e regressão. O *MAE* atingido foi de 5,25 e 5,30 anos para mulheres e homens, respectivamente, utilizando AAM para a representação facial. A ideia do LARR é primeiro realizar a regressão através do SVR para obter uma estimativa grosseira da idade e então determinar um ajuste local dentro de uma faixa etária restrita centrada nos resultados da regressão. O principal problema é como verificar diferentes idades dentro de uma faixa específica para o propósito de ajuste local, sendo que o objetivo é arrastar a idade inicialmente estimada pelo regressor global na direção mais próxima possível da idade verdadeira. Então, uma abordagem de classificação SVM para ajustar localmente ou verificar diferentes idades considerando cada label da idade como uma classe é utilizada, que melhorou significativamente o desempenho do estimador etário. Porém, o método LARR não pode determinar os parâmetros automaticamente dos classificadores para as faixas restritas da idade. Diferentes faixas são tentadas heurísticamente como 4, 8, 16, 32 e 64 e requer que o usuário escolha empiricamente a melhor solução. Para determinar os parâmetros automaticamente Guo et al. propuseram uma abordagem probabilística para combinar classificação e regressão. A ideia é transformar ou interpretar os resultados da regressão em probabilidades utilizando uma distribuição uniforme e então as probabilidades da classificação são cortadas pela distribuição uniforme. O desempenho foi melhorado, o *MAE* atingido foi de 5,11 e 5,12 anos para mulheres e homens da base YGA e 4,97 anos para a base de dados FG-NET.

## 2.3 Análise e síntese

A Tabela 2 mostra uma síntese e comparação de algumas técnicas existentes de estimação etária avaliadas em diversos conjuntos de dados. Nota-se que grandes esforços e estudos abrangentes têm sido mostrados na última década, seja para a representação do envelhecimento facial quanto para o desenvolvimento de algoritmos para a estimação etária. As comparações entre os diferentes métodos mostrados na tabela através de curvas *CS* podem ser recuperadas a partir de [21] e [25]. A partir da coluna *Banco de dados* da

Tabela 2 é possível observar que a base de dados FG-NET é utilizada em diversos trabalhos como benchmark para comparação com outras técnicas de estimação etária, como por exemplo, o método AGES. Também, nota-se que os métodos tiveram seus desempenhos avaliados a partir do *MAE* e *CS*, sendo tais métricas utilizadas neste trabalho.

A representação do envelhecimento facial e algoritmos para a estimação etária são duas questões essenciais para a estimativa etária a partir de imagens faciais. O modelo antropométrico e modelo de aparência ativa fornecem dados paramétricos para a representação facial. Porém, o modelo antropométrico tem um foco maior em mudanças no formato facial o qual é importante para a estimativa da idade de faces jovens. O AAM pode lidar com a estimativa da idade de pessoas jovens e idosas uma vez que possui informações da textura da face. Quando uma sequência de imagens faciais de um mesmo indivíduo em diferentes idades são disponíveis o método AGES pode se aplicado para capturar padrões no envelhecimento do indivíduo. *Age manifold* é muito útil quando uma grande base de dados com uma grande faixa etária está disponível, sendo que a extração de características e a redução da dimensionalidade estão interligadas. Os modelos de aparência são uma maneira de considerar casos quando múltiplas características faciais são mescladas como no caso de características globais e locais. A estimação etária pode ser abordada como um problema de classificação ou regressão uma vez que diferentes bases de dados e sistemas podem ser muito tendenciosa ou desbalanceada para a avaliação. Porém, não é puramente um problema de classificação ou regressão. Uma abordagem promissora é a combinação de métodos de classificação e regressão como demonstrados na subseção 2.2.3, sendo interessante o desenvolvimento de métodos para a combinação de classificadores e regressores para que o desempenho dos sistemas de estimação etária seja melhorado.

A partir da análise do estado da arte é possível verificar que os modelos de aparência foram pouco explorados assim como a análise do desempenho em conjuntos de dados públicos com grandes quantidades de imagens faciais, como no caso da base de dados MORPH. Portanto, neste trabalho será explorado novas técnicas de representação do envelhecimento facial através de modelos de aparência combinando características globais e locais de um indivíduo. Assim como a avaliação do método proposto no conjunto de dados de benchmark FG-NET e em bases de dados maiores como a MORPH, pouco explorada em trabalhos anteriores. As abordagens híbridas, pouco exploradas em trabalhos anteriores, serão investigadas neste trabalho, pois elas têm demonstrado um desempenho superior aos métodos de classificação e regressão quando considerados isoladamente.

No próximo capítulo será apresentada uma fundamentação teórica com conceitos fundamentais para o desenvolvimento deste trabalho. Serão apresentados algumas das principais abordagens presentes na literatura para a extração de características da textura de imagens faciais que serão analisada durante o desenvolvimento desta pesquisa. Também, serão apresentados os principais fundamentos de técnicas de aprendizagem de máquina como máquina de vetores de suporte e máquina de vetores de suporte para regressão nos quais serão utilizado para uma abordagem híbrida de classificação e regressão para a estimação etária.

Tabela 2: Síntese e comparação de técnicas de estimação etária. Utilizada a tabela de Fu et al. [2] como referência.

Publicação	Representação facial	Banco de dados (#indivíduos, #imagens)	Faixa etária	Algoritmo	Protocolo	MAE (anos)	CS ( $\leq 10$ )
Kwon & Lobo 1999	MA	Privado (N/A, 47)	Três grupos	C	15 imagens para teste	N/A	N/A
Kanno et al. 2001	MAP (mosaic)	Privado (110(M), 440(M))	12,15, 18,22	C (ANN)	N/A	N/A	N/A
Lanitis et al. 2004	AAM	Privado (40, 400)	0-35 anos	C ou R	50% trein. 50% teste	3,82 - 5,58	N/A
Geng et al. 2007	AGES	MORPH (515, 1430(M)+294(F))	15-64 anos	R	Trein. na FG-NET	8,83	70%
Zhou et al. 2005	MAP	FG-NET (82, 1002)	0-69 anos	R	5-fold cross val.	5,81	N/A
Geng et al. 2006	AGES	FG-NET (82, 1002)	0-69 anos	R	Leave one person out	6,77	81%
Geng et al. 2006	AGES	FG-NET (82, 1002)	0-69 anos	C	14 faixas, 5 anos cada	1,26	N/A
Yan et al. 2007	AAM	FG-NET (82, 1002)	0-69 anos	R (ranking)	Leave one person out	5,33	84%
Yan et al. 2007	AAM	FG-NET (82, 1002)	0-69 anos	R (RUN)	Leave one person out	5,78	84%
Yan et al. 2008	MAP (patches)	FG-NET (82, 1002)	0-69 anos	R (kernel reg.)	Leave one person out	4,95	N/A
Guo et al. 2008	AAM	FG-NET (82, 1002)	0-69 anos	R (LARR)	Leave one person out	5,07	88%
Guo et al. 2008	AAM	FG-NET (82, 1002)	0-69 anos	H	Leave one person out	4,97	88%
Yan et al. 2009	AAM	FG-NET (82, 1002)	0-69 anos	C (SSE)	Leave one person out	5,21	N/A
Xiao et al. 2009	AAM	FG-NET (82, 1002)	0-69 anos	C (mkNN)	300 trein. 702 teste	5,04	84%
Guo et al. 2009	AMF (BIF)	FG-NET (82, 1002)	0-69 anos	R (SVR)	Leave one person out	4,77	89%
Fu et al. 2007	AMF (OLPP)	YGA (800(M)+800(F), 4000(M)+4000(F))	0-93 anos	R (quadratic)	50% trein. 50% teste	8,0(M) 7,8(F)	70%(M) 70%(F)
Fu & Hung, 2008	AMF (CEA)	YGA (800(M)+800(F), 4000(M)+4000(F))	0-93 anos	R (quadratic)	50% trein. 50% teste	6,0(M) 5,5(F)	82%(M) 83%(F)
Yan et al. 2008	MAP (patches)	YGA (800(M)+800(F), 4000(M)+4000(F))	0-93 anos	R (kernel reg.)	1000 trein. 3000 teste	4,38(M) 4,94(F)	88%(M) 85%(F)
Guo et al. 2008	AMF (OLPP)	YGA (800(M)+800(F), 4000(M)+4000(F))	0-93 anos	R (LARR)	4-fold cross val.	5,30(M) 5,25(F)	83%(M) 81%(F)
Guo et al. 2008	AMF (OLPP)	YGA (800(M)+800(F), 4000(M)+4000(F))	0-93 anos	H	4-fold cross val.	5,12(M) 5,11(F)	83%(M) 82%(F)
Zhuang et al. 2008	MAP (patches)	YGA (800(M)+800(F), 4000(M)+4000(F))	0-93 anos	R (HMM)	50% trein. 50% teste	5,40(M) 6,33(F)	82%(M) 76%(F)
Guo et al. 2009	MAP (BIF)	YGA (800(M)+800(F), 4000(M)+4000(F))	0-93 anos	C (SVM)	4-fold cross val.	3,47(M) 3,91(F)	88%(M) 85%(F)
Guo et al. 2009	MAP (BIF) + AMF	YGA (800(M)+800(F), 4000(M)+4000(F))	0-93 anos	C (SVM)	4-fold cross val.	2,58(M) 2,61(F)	N/A
Choi et al. 2011	AAM + LBP + GW	FG-NET (82, 1002)	0-69 anos	C+R (SVM+SVR)	Leave one person out	4,65	87%
Kohli et al. 2013	AAM	FG-NET (82, 1002)	0-69 anos	C+R (SVM+SVR)	802 train 200 test	3,85	83%

Observação: N/A-não disponível, M-Masculino e F-Feminino. Na coluna "Representação facial", MA-Modelo Antropométrico, AAM-Active Appearance Model, AGES-Aging Pattern Subspace, AMF-Age Manifold, e MAP-Modelo de Aparência. Na coluna "Algoritmo", C-Classificação, R-Regressão, e H-Híbrido. Na coluna "CS"o erro absoluto tolerável é  $\pm 10$  anos.

## 3 Fundamentação teórica

Neste capítulo serão apresentados métodos de extração de características faciais os quais são fundamentais para o desenvolvimento de sistemas de reconhecimento de padrões, como no caso da estimação etária abordada neste trabalho. Primeiramente, um breve estudo sobre o envelhecimento facial é apresentado de forma que seja possível definir as características relacionadas ao envelhecimento como rugas, manchas e formato facial. Adicionalmente, são apresentados métodos para a classificação e regressão que serão utilizados para a estimativa da idade.

### 3.1 Um breve estudo sobre envelhecimento

Segundo Stone [8] o processo de envelhecimento é lento, severo e irreversível. E ocorre em diferentes taxas de indivíduo para indivíduo e também em cada pessoa em qualquer momento. O envelhecimento facial não está relacionado com a atividade física em geral a menos que se esteja muito doente e acamado. As mudanças gerais que ocorrem com o envelhecimento podem ser:

- Reabsorção óssea ao longo da borda orbital inferior (borda óssea embaixo do olho) e na região da junção alar/nariz bochecha. O restante dos ossos do maxilar superior e inferior não se alteram e os dentes permanecem intactos;
- Atrofias faciais;
- A elasticidade da pele diminui e a pele é afinada.

Ao longo do tempo os efeitos da gravidade e a tração longitudinal dos músculo causam a caída ou flacidez da pele e de estruturas mais profundas das bochechas, pálpebras, nariz, queixo etc., causando as rugas e papadas na face. Rugas devido ao movimento muscular são referidas como rugas dinâmicas (conhecidas como rugas de ex-

pressão) e aquelas que ocorrem devido ao processo de envelhecimento são definidas como rugas estáticas.

Na faixa etária de 30 a 40 anos as pálpebras superiores começam a cair e pés de galinha aparecem nos cantos externos dos olhos. Stone diz que os sulcos nasolabiais aprofundam-se e também 40% dos homens tem algum grau de perda de cabelos. Na faixa etária de 40 a 50 anos as pálpebras caem ainda mais, o espaço compreendido entre as sobrancelhas, chamado glabella, fica enrugado e rugas na testa aparecem e 25% dos homens são calvos. As sobrancelhas exteriores começam a cair. As gorduras das bochechas começam a descer em direção a linha do queixo. Excesso de pele no pescoço aparecem. Em alguns indivíduos os olhos se tornam mais fundos devido ao atrofiamento de gordura ao invés de formar bolsas nas pálpebras. Na faixa etária de 50 a 60 anos as rugas no pescoço, papadas e linhas de marionete começam a aparecer. As sobrancelhas caem mais assim como a ponta do nariz. Os lábios afinam mostrando pouca área avermelhada e aparecem rugas nessa região. Na faixa etária de 60 a 70 anos a pele facial afina, ocorre o atrofiamento da gordura na região das bochechas e têmpora e as células de pigmento da pele aumentam em número e tamanho, causando manchas marrons nas costas das mãos e face. E 50% dos homens dessa faixa são carecas. Na Figura 8 são mostrados alguns esboços de envelhecimento facial de 30 a 80 anos com 10 anos por esboço.

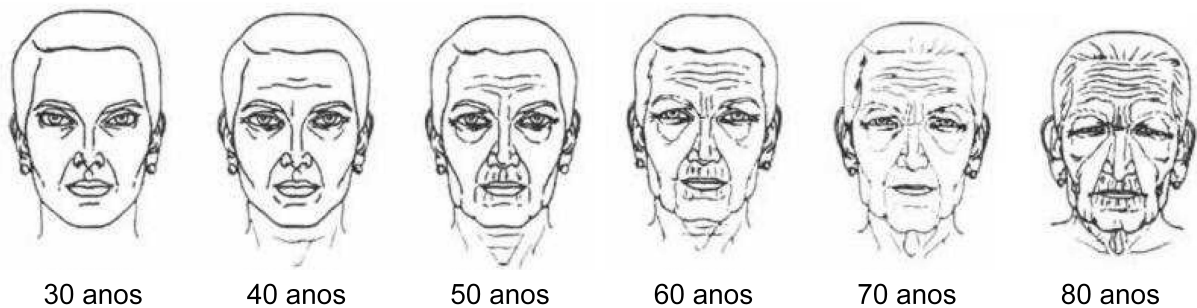


Figura 8: Esboços de envelhecimento facial de 30 a 80 anos com 10 anos por esboço [8].

Durante o processo de envelhecimento, Stone diz que o contorno geral da face muda de um triângulo com o vértice apontando para baixo ou contorno em forma de "U" para um trapézio ou retângulo devido a flacidez da pele e a caída dos tecidos moles da face, conforme ilustrado na Figura 9. E também que essa mudança geométrica é mais evidente no rosto feminino. O processo de envelhecimento é acelerado pelos seguintes fatores: tabagismo, predisposição genética, estresse emocional, doenças, mudanças drásticas no peso e exposição a climas extremos.

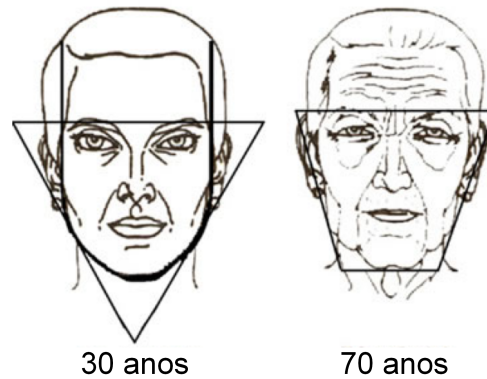


Figura 9: Mudanças no contorno facial devido ao envelhecimento [8].

## 3.2 Extração de características

### 3.2.1 Modelos de aparência ativa

#### 3.2.1.1 Modelagem estatística

Uma variável aleatória  $x$  é uma função que atribui um resultado numérico a cada resultado individual de uma experiência. Se analisarmos os seres humanos verificamos que suas faces são formadas por elementos básicos como olhos, nariz, boca, rugas etc., independente de sua configuração. Ou seja, identificamos olhos em todos os humanos apesar de cada indivíduo apresentar um formato e textura distintos para os olhos, por exemplo. Portanto, cada elemento da face pode ser encarado como uma variável aleatória e, por consequência, a própria face é uma variável aleatória. Dessa forma, é possível trabalhar com um conjunto discreto de valores e modelar estatisticamente as faces.

O modelo de aparência ativa (AAM) [9] é um algoritmo capaz de modelar estatisticamente um objeto baseado na decomposição de forma e textura que compõem um conjunto de amostras de treino similares submetidos a um treinamento. Além de modelar uma face o AAM aplica PCA para a redução de dimensionalidade, diminuindo o custo computacional de operações que necessitem representações da face em etapas futuras. Para aplicar o AAM é necessário um conjunto de amostras que representem a variável aleatória a ser modelada. No caso de faces o conjunto de treinamento é formado por imagens faciais. A delimitação da região da face e plano de fundo é realizada por um conjunto de pontos de referência que constituem o contorno da região que contém os componentes de interesse da face, englobando olhos, nariz, boca, sobrancelhas e o próprio formato da face. Esse de pontos de referência são conhecidos como *landmarks* e todas as imagens do

banco de dados devem possuir esses pontos extraídos. Com a delimitação da região da face e a obtenção dos *landmarks*, é possível gerar uma distribuição espacial para a variável aleatória obtida com os pontos do banco de dados e gerar um modelo de forma para a face. Com os mesmos pontos é possível mapear regiões, através de triangulação e utilizar os pixels delimitados por essas regiões como outra variável aleatória e gerar um modelo de textura. A Figura 10 ilustra as *landmarks* em uma imagem facial, assim como o formato livre da aparência.

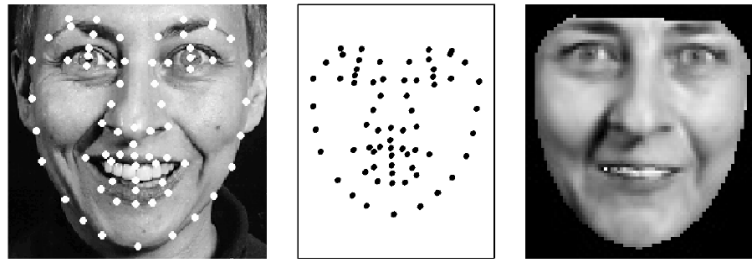


Figura 10: Imagem facial com os pontos de referência, conjunto de pontos e o formato livre da imagem, respectivamente [9].

### 3.2.1.2 Modelo estatístico da forma

O primeiro modelo gerado pelo AAM é o modelo de forma (*shape model*). Devido não existir uma padronização nas imagens de entrada em relação à posição ou inclinação da face, as  $s$  imagens dos dados de treinamento podem estar em escalas e orientações diferentes, sendo necessário um procedimento de alinhamento dos pontos entre cada forma (*landmark*)  $x$  dos  $s$  exemplos de treinamento para gerar o modelo para cada imagem  $i$ . Como cada ponto é formado por duas coordenadas, têm-se um conjunto de  $2n$  pontos, onde  $n$  é o número de *landmarks*. Considerando que o formato de um objeto é invariante a transformações Euclidianas, pode-se alinhar cada vetor de treinamento através de translações, rotações e escalas. É realizada a análise de Procrustes para reduzir os efeitos de escala, rotação e translação com os seguintes passos:

1. aplica-se translação em cada exemplo de treinamento de forma que seu centro de massa coincida com a origem;
2. define-se um dos exemplos de treinamento como sendo uma estimativa inicial da forma média e normaliza-se o vetor de forma que  $\|\bar{x}\| = 1$ ;
3. alinha-se todos os exemplos de treinamento de acordo com a estimativa para forma média;



4. incrementa-se  $i$  e estima-se um novo vetor de forma média  $\|\bar{x}^{i+1}\|$  utilizando os exemplos alinhados no passo anterior. Normaliza-se  $\|\bar{x}^{i+1}\|$ ;
5. avalia-se a equação de custo  $D = \sum_{i=1}^s \|x^i - \bar{x}\|$ , objetivando minimizar a distância entre cada exemplo de treinamento do vetor de forma média  $\bar{x}$ . Verifica-se se houve convergência. Caso positivo, a análise termina. Caso contrário, retorna-se ao terceiro passo até obter a convergência.

O procedimento acima fornece uma medida estatística da distribuição de cada ponto de entrada. Então, é necessário equacionar um modelo  $M$  com parâmetros  $b$  que represente a diferença entre os exemplos de treinamento e a forma média de treinamento na forma  $x = M(b)$ . O vetor de forma é definido como  $b$ . A dimensão máxima de  $b$  será  $2n$ , utilizando todos os *landmarks*. Para utilizar um modelo contendo apenas componentes significativas, é necessário a utilização de PCA para a redução de dimensionalidade para que o parâmetro  $b$  assumira uma dimensão  $b \leq 2n$ . Para a construção do modelo, computa-se o vetor de forma média, definido na Equação 3.1.

$$\bar{x} = \frac{1}{s} \sum_{i=1}^s x^i \quad (3.1)$$

É calculada a covariância amostral conforme Equação 3.2.

$$S = \frac{1}{s-1} \sum_{i=1}^s (x^i - \bar{x})(x^i - \bar{x})^T \quad (3.2)$$

Então, é calculado os autovetores (*eigenvectors*)  $\Phi_i$  e os autovalores (*eigenvalues*)  $\lambda_{s,i}$  associados à  $S$ . Os autovalores são ordenados de forma decrescente. Considerando os  $z$  autovetores associados aos  $z$  maiores autovalores, para  $z \leq 2n$ , pode-se reduzir a dimensionalidade e aproximar qualquer exemplo como

$$x \approx \bar{x} + P_s b_s \quad (3.3)$$

onde  $P_s = [\Phi_1 | \Phi_2 | \dots | \Phi_z]$  e  $b_s$  é o parâmetro de forma e representa  $b$  após a redução de dimensionalidade. Ou seja,  $P_s$  é o novo espaço formado pelos principais autovetores dos dados de forma e  $b$  é responsável pelo peso atribuído a cada autovetor, gerando diferentes formas no novo espaço.

Uma forma qualquer pode então ser representada pela forma média mais uma deformação gerada pela associação dos autovetores ponderados pelo parâmetro  $b_s$ . Os parâmetros do modelo  $M$  podem ser definidos através da matriz  $P_s$  e a forma média de acordo com a seguinte Equação 3.4.

$$b_s^i = P_s^T(x^i - \bar{x}) \quad (3.4)$$

Essa etapa é denominada *Active Shape Model* (ASM) [9], porém, mais informações podem ser obtidas a partir da textura da imagem que contém também informações importantes relacionada ao envelhecimento.

### 3.2.1.3 Modelo estatístico de textura

O modelo estatístico de textura deve ser independente das formas das amostras de treinamento. Então, é realizada uma deformação em cada imagem de maneira que seus pontos referentes à forma coincidam com o modelo de forma média. No caso bidimensional pode-se utilizar a triangulação de Delaunay para delimitar as áreas que compõem a textura da face utilizando os *landmarks* como referência. Uma transformação em cada amostra de treinamento onde cada pixel original no interior do triângulo formado por 3 pontos que compõem a forma original é mapeado em um novo pixel em uma imagem deformada para a forma média de acordo com a posição de cada vértice dos triângulos. Ou seja, para o modelo de textura a informação de formato deve ser desconsiderada fazendo um remapeamento através dos pontos delimitados por Delaunay onde cada face de entrada é levada à forma média.

As imagens de treinamento podem possuir diferentes condições de iluminação, interferindo na análise estatística da textura. Por exemplo, regiões com reflexos ou pouco iluminadas, distorcem os valores reais dos pixels que compõem a imagem. Portanto, há a necessidade de se alinhar fotometricamente as texturas de treinamento para minimizar os efeitos indesejados introduzidos pelas diferentes condições de iluminação de cada imagem de entrada. O processo de alinhamento é realizado normalizando as texturas conforme a Equação 3.5 definida abaixo

$$u : g^i = \frac{g^i - g^{-i}}{\sigma^i} \quad (3.5)$$

onde  $g^i$  e  $g^{-i}$  são a intensidade e a média das intensidades dos pixels da textura, respec-

tivamente, e  $\sigma^i$  é o desvio padrão para as intensidades dos pixels da imagem de entrada  $i$ . Similar ao modelo de forma, a distância Euclidiana  $E_g$  entre cada vetor de textura de treino e a textura média é minimizada segundo a Equação 3.6 definida abaixo

$$E_g = \sum_{i=1}^s \|g^i - \bar{g}\| \quad (3.6)$$

onde  $\bar{g}$  é a textura média para o conjunto de treinamento.

O modelo de textura é inicializado utilizando uma das amostras de treinamento como referência. Através de iterações sucessivas controladas pelo índice de cada instância  $i$ , nesse modelo também deseja-se minimizar  $E_g$ , para obter um vetor de textura  $\bar{g}$  médio, após convergência.

Aplicando PCA, pode-se representar a textura em função de parâmetros de textura  $b_g$  como  $g \approx \bar{g} + P_g b_g$ , onde  $P_g$  contém os  $k$  autovetores correspondentes aos maiores  $k$  autovalores  $\lambda_g$ . Portanto, considerando os dois modelos, pode-se definir uma imagem de entrada em termos de  $b_s$  e  $b_g$ . O modelo de textura apresenta  $\bar{g}$  e a matriz  $P_g$ , conforme Equação 3.7 descrita abaixo

$$b_g = P_g^T (g - \bar{g}) \quad (3.7)$$

A textura de qualquer imagem no formato médio pode ser representada através da textura média do modelo somada aos autovetores ponderados pelo vetor de textura  $b_g$ .

#### 3.2.1.4 Modelo estatístico combinado

Modelos estatísticos de forma e textura não devem ser independentes, pelo fato de que para alterar a textura de uma face é necessário alterar também parâmetros locais relacionados a posição de componentes faciais, portanto, é necessário alterar a forma ou posicionamento dos *landmarks* também. Assim como alterar o formato de uma face introduz mudanças nas texturas. O modelo que junta informações de forma e textura, minimizando redundância de informação é definido como modelo combinado ou modelo de aparência, denominado por  $c$ .

O modelo de aparência utiliza o parâmetro de forma  $b_s$  e o parâmetro de textura  $b_g$ , combinando as informações em um único parâmetro de aparência  $b_{sg}$ . Porém,  $b_s$  está na forma de unidades de distância, ao qual refere-se ao posicionamento do pixel, já o

parâmetro  $b_g$  possui unidade de intensidade de pixel. A relação entre os dois parâmetros é realizada atribuindo uma matriz de pesos  $W_s$  ao parâmetro de forma, relacionando o efeito que alterar  $b_s$  gera nas intensidades de pixels relacionadas à  $b_g$ , conforme mostra a Equação 3.11 definida abaixo

$$b_{sg} = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} \quad (3.8)$$

onde  $W_s$  é calculado variando-se a forma de um exemplo do banco de dados através de um ajuste no valor de  $b_s^i$  e observando o efeito resultante na intensidade dos pixels que são refletidos na textura e, conseqüentemente, nos valores de  $b_g^i$ . Uma média obtida para cada um dos  $W_s^i$  resultantes é utilizada como matriz de peso, definida pela Equação 3.9

$$W_s = \frac{1}{N} \sum_{i=0}^N W_s^i \quad (3.9)$$

$W_s$  é uma matriz diagonal, onde cada elemento relaciona os pesos atribuídos aos autovetores de  $P_s$  que devem ser alinhados com  $P_g$  no modelo combinado. Outra possibilidade para estimar  $W_s$  com menor custo computacional é definida pela matriz:

$$W_s = \begin{bmatrix} W_{11} & 0 & \dots & 0 \\ 0 & W_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_{z,k} \end{bmatrix}, \quad (3.10)$$

onde

$$W_{ij} = \frac{\sum_{i=1}^z \lambda_{s,i}}{\sum_{i=1}^k \lambda_{g,j}} \quad (3.11)$$

O modelo final é obtido aplicando-se PCA em  $b_{sg}$ ,  $b_{sg} \approx P_c c$ , onde  $P_c$  é a matriz de autovetores associada aos primeiros  $m$  maiores autovalores. Dessa forma, tem-se uma redução de dimensionalidade  $m \leq z + k$ , onde  $z$  e  $k$  estão associados aos números de autovetores obtidos nos modelos de forma e textura, respectivamente.

A matriz  $P_c$  é obtida aplicando-se PCA na matriz formada pelos parâmetros de textura  $b_g$  e pelos parâmetros de forma ponderados  $W_s b_s$ . Portanto, resultam uma parcela

relacionada à textura  $P_{cg}$  e outra relacionada à forma  $P_{cs}$ , de acordo com a Equação 3.12.

$$P_c = \begin{pmatrix} P_{cs} \\ P_{cg} \end{pmatrix} \quad (3.12)$$

Qualquer face pode ser representada por uma alteração de forma e textura, alterando-se  $b_s$  e  $b_g$  que, no modelo combinado, é representado, pelo vetor de aparência  $c$ . No entanto, qualquer face pode ser representada pelo modelo combinado e um vetor de aparência. Como o modelo é único, qualquer face é representada pelo vetor de aparência. Uma imagem então pode ser modelada em termos de seu formato e textura através das seguintes relações:

$$x = \bar{x} + P_s W_s^{-1} P_{cs} c \quad (3.13)$$

e

$$g = \bar{g} + P_g P_{cg} c. \quad (3.14)$$

### 3.2.2 *Gabor wavelets*

Gabor *wavelets* é um conjunto de filtros gerados a partir de uma mesma função Gabor com a variação de alguns parâmetros. São similares ao campo visual receptivo 2D das células corticais dos mamíferos mostrando fortes características de localização espacial e seletividade de orientação para representar a informação necessária a ser extraída em forma de orientações e frequências. Possui grande relevância biológica e propriedades computacionais [30].

Introduzidos somente na última década [31], *wavelets* são uma abordagem relativamente recente para o processamento de sinais. A sua principal vantagem é permitir uma análise multiresolução, ou seja, em diferentes escalas ou resoluções. Também, permitem decimação simultânea no espaço e na frequência. As transformadas mais clássicas permitem apenas a decimação em frequência na transformada em si e no tempo ou posição na transformada inversa. Dessa maneira, a transformada de Fourier tem como resultado o conteúdo frequencial de toda uma imagem, ou seja, a contribuição da imagem em um determinado componente de frequência. A decimação simultânea permite descrever uma imagem em termos de frequências que ocorrem em uma posição, em oposição a uma ca-

pacidade para medir o conteúdo de frequência em toda a imagem. Claramente, isso dá um poder descritivo maior.

Primeiramente, é necessário uma função de base para poder decompor um sinal. A transformada de Fourier possui uma função de base senoidal em diferentes frequências. A função da transformada de Fourier é realizar uma convolução dessas senóides com um sinal para determinar o quanto de cada uma dessas senóides está presente no sinal. Gabor *wavelet* é bem adequado para fins de introdução, uma vez que é essencialmente uma onda senoidal modulada por um envelope Gaussiano. O Gabor *wavelet*,  $gw$ , é definido como

$$gw(t, \omega_0, t_0, a) = e^{-j\omega_0 t} e^{-\left(\frac{t-t_0}{a}\right)^2} \quad (3.15)$$

onde  $\omega_0 = 2\pi f_0$  é a frequência modulada,  $t_0$  a posição, e  $a$  controla a largura do envelope Gaussiano, que envelopa o sinal oscilante. Um exemplo de Gabor *wavelet* é mostrado na Figura 11 a qual mostra a parte real e imaginária, sendo o módulo o envelope Gaussiano. Aumentando o valor de  $\omega_0$  aumenta-se a frequência dentro do envelope, enquanto aumentando o valor de  $a$  estende-se o envelope sem afetar a frequência. Se o lado direito da Equação 3.15 for removido, tem-se uma função de base senoidal da transformada de Fourier, ou então, colocando um valor muito grande para  $a$ . Alternativamente, colocando o valor zero em  $f_0$ , a informação de frequência é removida. Operando entre esses extremos, é possível obter a informação de posição e frequência simultaneamente.

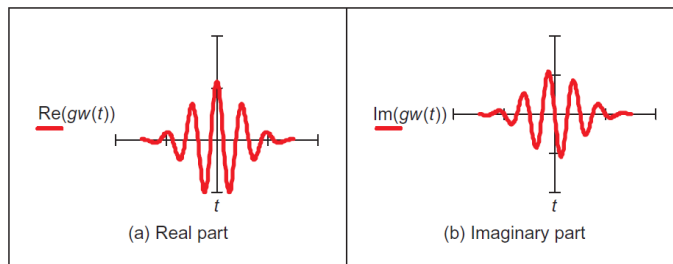


Figura 11: Exemplo de Gabor *wavelet* [10].

A função Gabor bidimensional tem como objetivo ser o ideal em termos de resolução espacial e espectral. Esses Gabor *wavelets* 2D são definidos por

$$gw2D(x, y) = \frac{1}{\sigma\sqrt{\pi}} e^{-\left(\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}\right)^2} e^{-j2\pi f_0((x-x_0)\cos(\theta)+(y-y_0)\sin(\theta))} \quad (3.16)$$

onde  $x_0$  e  $y_0$  controlam a posição,  $f_0$  a frequência de modulação ao longo de cada eixo, e

$\theta$  controla a orientação do *wavelet*.

Um exemplo de Gabor *wavelets* 2D é mostrado na Figura 12, o qual mostra que a parte real é uma função par e a parte imaginária ímpar. Novamente, diferentes valores para  $f_0$  e  $\sigma$  controlam a frequência e o envelope respectivamente, o parâmetro  $\theta$  controla a rotação. A função de uma transformada *wavelet* é determinar onde e como cada *wavelet* especificado pelo intervalo de valores para cada um dos parâmetros ocorrem em uma imagem. A Figura 13 mostra um exemplo da transformada, onde os parâmetros foram escolhidos de maneira a selecionar características faciais, como olhos, nariz e boca [10]. Essas características estão onde há conteúdo de frequência local com a orientação de acordo com a inclinação da cabeça. Isso mostra a capacidade do Gabor *wavelet* para selecionar e analisar variações localizadas na intensidade da imagem.

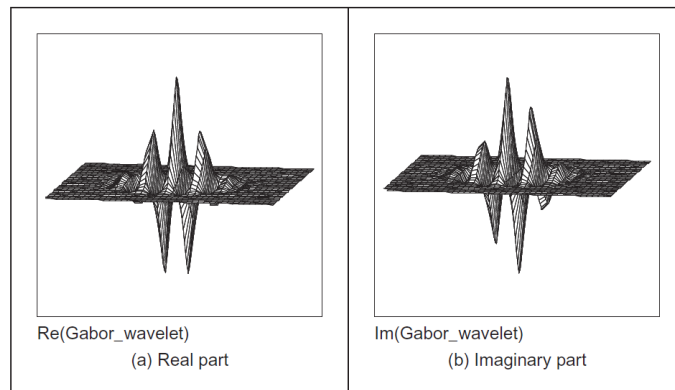


Figura 12: Exemplo Gabor *wavelet* 2D [10].

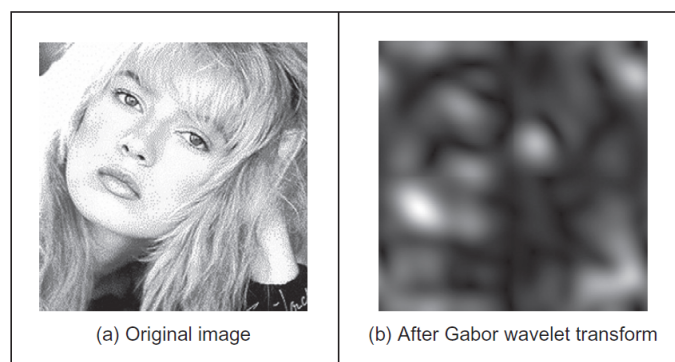


Figura 13: Exemplo da transformada Gabor wavelet [10].

A teoria de Gabor *wavelets* é naturalmente muito poderosa, uma vez que acomoda frequência e posição simultaneamente, facilitando a análise em multiresolução. A análise é sensível a escala, sendo muito vantajosa, pois imagens que estão longe da câmera parecem ser menores que aquelas que estão mais próximas. Dentre as diversas aplicações de Gabor

*wavelets* estão a análise da textura da íris para ser utilizada em sistemas biométricos de segurança, extração de características faciais para sistemas de reconhecimento facial automático, restauração de imagens, análises de texturas etc.

### 3.2.3 *Local binary patterns*

*Local binary pattern* (LBP) é um método para descrição de textura que tem recebido muita atenção da comunidade científica devido seu atrativo desempenho em diversas aplicações de visão computacional, como por exemplo, reconhecimento facial e biometria. A abordagem inicial originou na Universidade de Oulu [32] e tem sido refinada ao longo do tempo até a abordagem mais recente [33]. Essencialmente, o LBP básico de uma região  $3 \times 3$  é derivado da comparação do ponto central com seus vizinhos para então ser gerado o código que irá substituir o valor do ponto central. Para os pontos  $P$  e  $\mathbf{P}_x$  o processo depende de um limiar, de acordo com a função

$$s(x) = \begin{cases} 1, & \text{se } \mathbf{P}_x > P \\ 0, & \text{caso contrário} \end{cases} \quad (3.17)$$

O código é derivado de um ponderação binária aplicada ao resultado da limiarização, sendo equivalente a limiarizar os pontos vizinhos do ponto central e em seguida gerar um código em formato binário. Portanto, o código LBP para um ponto  $P$  com oito vizinhos  $x$  é definido como

$$LBP = \sum_{x \in \{1,8\}} s(x) 2^{x-1} \quad (3.18)$$

Na Figura 14 é exemplificado o processo LBP, onde o ponto LBP é o ponto central e os oito valores para  $x$  seus vizinhos imediatos. Para o caso de uma região  $3 \times 3$  na Figura 14(a), o valor do ponto central é excedido três vezes, portanto existem três 1s no código resultando na Figura 14(b). Quando desembrulhado no sentido horário, considerando o ponto superior central o bit mais significativo, o código resultante é 101000012. Quando considerado como um código binário com ponderações mostrados na Figura 14(c), é possível encontrar o valor final  $LBP = 161$  (Figura 14(d)). Naturalmente, o processo de limiarização, desembrulhamento e ponderações podem ser realizadas de diferentes maneiras, porém é essencial que sejam consistentes em toda a imagem. O código  $P$  codifica a estrutura de intensidade local, ou padrão binário local.



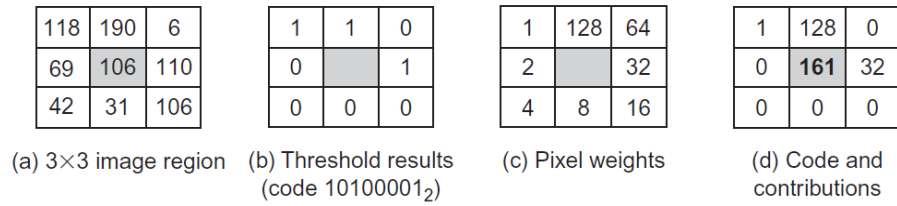


Figura 14: Processo de construção de um código LBP [10].

A abordagem LBP determina um histograma dos códigos derivados de uma imagem inteira, descrevendo a textura. A técnica é inerentemente invariante a translação pela sua formulação, ou seja, uma textura deslocada deve possuir o mesmo histograma com os códigos LBP. Porém, o processo básico não é invariante a escala e rotação, no caso da rotação, uma diferente ponderação será aplicada aos pontos resultando em um código diferente.

O histograma dos valores LBP da textura de uma tela francesa (*French canvas*) (Figura 15(a)) é mostrado na Figura 15(b). O histograma é também mostrado para a mesma textura deslocada para a esquerda (Figura 15(c)) em 40 pixels, com seu histograma resultante na Figura 15(d), sendo muito similar, como esperado. De fato existem algumas diferenças entre os pontos entre os dois histogramas, uma vez que foi realizado utilizando imagens reais, mas a diferença é de aproximadamente 20 pixels, sendo uma diferença muito pequena considerando que alguns valores do histograma são contados em mais de  $4 \times 10^3$  pixels e não podem ser determinadas por inspeção visual [10].

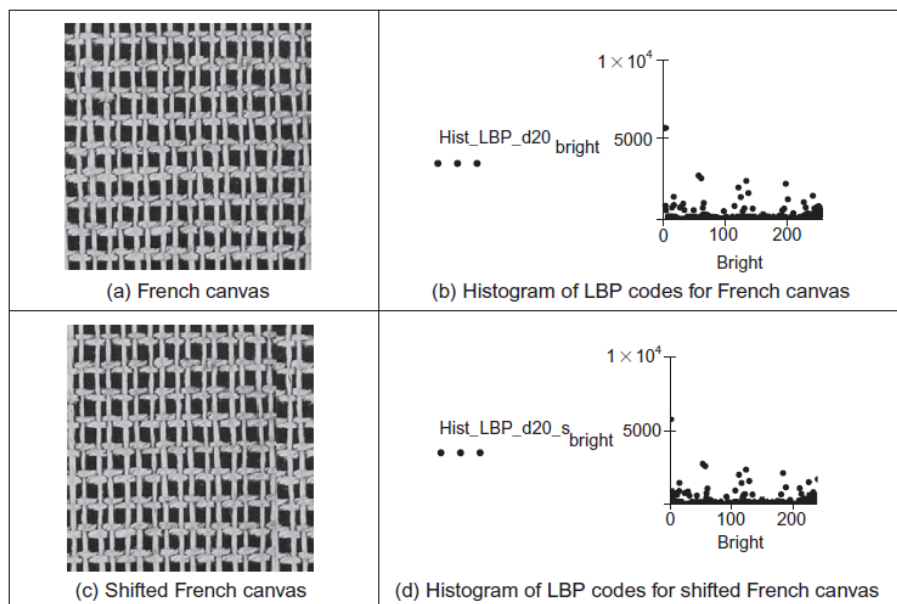


Figura 15: LBP invariante a translação [10].

Para que o LBP seja invariante a escala é necessário considerar pontos em uma maior distância. Se o espaço é amostrado de uma maneira circular, e  $P$  pontos são derivados de um raio  $R$ , então as equações de coordenadas para  $i \in (1, P)$  são

$$x(i) = \begin{bmatrix} x_0 + R \cos\left(\frac{2\pi}{P}i\right) \\ y_0 + R \sin\left(\frac{2\pi}{P}i\right) \end{bmatrix} \quad (3.19)$$

Como é possível ter diferente número de pontos no código, a geração do código para ser invariante a escala será

$$LBP_{S(P,R)} = \sum_{x \in 1,P} s(x(i))2^{i-1} \quad (3.20)$$

Os padrões para uma amostragem radial em diferentes valores de  $P$  e  $R$  são mostrados na Figura 16, onde a Figura 16(a) é a amostragem de um círculo com 8 pontos e raio 1, sendo equivalente ao padrão 3X3, a Figura 16(b) é para um raio de 2 com 8 pontos e a Figura 16(c) para um raio de 3 com 16 pontos. Todas mostram o efeito da discretização em baixa resolução de padrões circulares e é mais comum usar interpolação para determinar os valores dos pontos ao invés de utilizar o valor do pixel mais próximo.

$\mathbf{c} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$	$\mathbf{c} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$\mathbf{c} = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}$
(a) $\mathbf{c} := \text{Circle}(8,1)$	(b) $\mathbf{c} := \text{Circle}(8,2)$	(c) $\mathbf{c} := \text{Circle}(16,3)$

Figura 16: Padrões circulares para (P,R) [10].

Para que o LBP seja invariante a rotação é necessário deslocar o código derivado de modo a alcançar um número inteiro mínimo, sendo definido como

$$LBP_{R(P,R)} = \min \left\{ ROR \left( \sum_{x \in 1,P} s(x(i))2^{i-1} \right) \right\} \quad (3.21)$$

onde  $ROR()$  é o operador de rotação. A fim de atingir melhor capacidade discriminatória, observou-se que alguns padrões básicos predominaram na capacidade discriminatória e a ocorrência desses padrões predominaram na capacidade de descrição de textura [33].

Considerando o  $LBP_{S(8,1)}$  (Figura 16(a)) e que a saída limiarizada relativa ao ponto central seja preto ("0") ou branco ("1"), é possível encontrar as disposições dadas pela Figura 17. Os padrões 0-8 correspondem a características básicas, o padrão 0 representa o limiar para um ponto brilhante (todos os pontos circundantes tem um valor mais baixo) e o padrão 8 representa um ponto escuro (todos os pontos são mais brilhantes). Os padrões 1-7 representam linhas de vários graus de curvatura, o padrão 1 representa o término de uma linha, padrão 2 um pontiagudo e o padrão 4 representa uma borda. Esses são chamados de padrões binários uniformes (*uniform binary patterns*) e são caracterizados por ter no máximo duas transições de "1-0" (ou vice-versa) quando progredidos em torno do padrão circular.

Os padrões remanescentes, sem um label, possuem mais que duas transições de "1-0" e são chamados de não-uniformes (*nonuniform*). Esses padrões podem ocorrer em qualquer rotação, então o LBP pode ser organizado de forma a detectá-los para ser invariante a rotação.

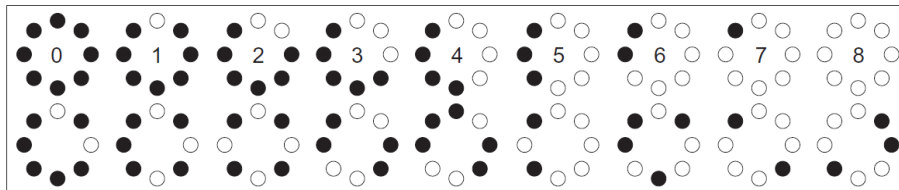


Figura 17: Padrões invariantes a rotação para  $LBP_{S(8,1)}$  [10].

Para a abordagem LBP uniforme, primeiramente é necessário detectar se os padrões são uniformes ou não, para isso é necessário utilizar o operador  $U$  que conta o número de transições de "1-0" (ou vice-versa):

$$U(LBP_{S(P,R)}) = |s(x(0)) - s(x(P0))| + \sum_{x \in 1, P-1} |s(x(i)) - s(x(i+1))| \quad (3.22)$$

Considerando o  $LBP_{S(8,1)}$ , o padrão 0-8 possui um valor máximo de  $U = 2$  ( $U = 0$  para os padrões 0 e 8 e  $U = 2$  para todos os outros). Então, é necessário determinar um código para cada padrão uniforme. Sendo que os padrões são invariantes a rotação, é possível determinar o código simplesmente contando o número de bits que estão ativos no padrão. Para os padrões não-uniformes, todos os bits são ajustados para terem o mesmo valor do código. Para os padrões definidos na Figura 17, existem códigos 0-8, então pode-se aglomerar os padrões não-uniformes para um código com valor 9. Para  $N$

padrões entre 0 e  $N - 1$ , o código invariante a rotação é definido como

$$LBP_{U(P,R)} = \begin{cases} \sum_{i \in 1,P} s(x(i)), & \text{se } U < 3 \\ N, & \text{caso contrário} \end{cases} \quad (3.23)$$

Com o histograma de ocorrências dessas características básicas é possível descrever a textura pela frequência da ocorrência dessas estruturas básicas locais.

Aplicando o descritor LBP uniforme na textura anterior (*French Canvas*) e na versão deslocada e rotacionada, o histograma dos códigos são similares. Sendo os códigos mais predominantes aqueles para estruturas de linhas, as quais são consistentes com a imagem utilizada. Os códigos 0-8 também dominam a representação, como esperado. Os resultados da versão deslocada e rotacionada são mostrados na Figura 18, com a versão deslocada na Figura 18(b) e a versão rotacionada na Figura 18(c). Visualmente, as diferenças são pequenas entre os histogramas na Figura 18(b) e (d), sendo as diferenças menores que 100, consideravelmente menor que  $10^4$  exposto pela técnica LBP uniforme [10].

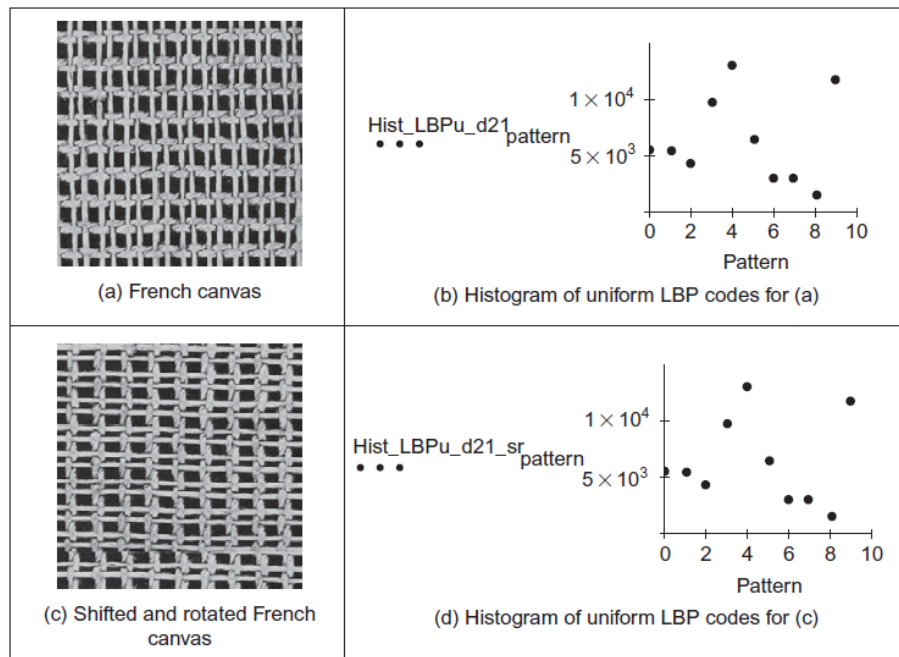


Figura 18: Histogramas de LBP uniformes [10].

### 3.2.4 *Local phase quantization*

Em visão computacional a análise de texturas é muito importante e tem diversas aplicações incluindo inspeção de superfícies, análise de imagens médicas, reconhecimento facial etc. Em algumas aplicações a degradação da imagem pode limitar a aplicabilidade das informações provenientes de texturas. Essas degradações podem acontecer quando a imagem é desfocada devido a movimentos, imagem está fora de foco ou devido a perturbações atmosféricas. Uma vez que desborrar uma imagem é computacionalmente caro é desejável analisar a textura de maneira que seja insensível ao borramento.

*Local phase quantization* (LPQ) é um método baseado na propriedade de invariância ao borramento do espectro de fase de Fourier para classificação de textura onde utiliza a informação de fase local extraída utilizando a transformada discreta 2D de Fourier (*Discrete Fourier Transform* - DFT) computada sobre uma vizinhança retangular (janela local) para cada pixel da imagem. A Figura 19 exemplifica algumas imagens faciais com variações na iluminação e borramentos.



Figura 19: Imagens faciais com variações na iluminação e borramentos [11].

O operador LPQ é aplicado para a descrição de textura, sendo processado localmente em cada pixel e os códigos resultantes são apresentados em forma de histograma, sendo similar ao método LBP [34]. Em processamento digital de sinais o modelo discreto para que o borramento de uma imagem original  $f(x)$  seja invariante no espaço resultando em uma imagem observada  $g(x)$  pode ser expressa pela seguinte convolução

$$g(x) = (f * h)(x) \quad (3.24)$$

onde  $h(x)$  é a função de dispersão do ponto (*Point Spread Functions* - PSF) de desfoque,  $*$  é a convolução 2D e  $x$  é o vetor de coordenadas  $[x, y]^T$ . No domínio Fourier é definido

como

$$G(u) = F(u)H(u) \quad (3.25)$$

onde  $G(u)$ ,  $F(u)$  e  $H(u)$  são as transformadas discretas de Fourier da imagem borrada  $g(x)$ , da imagem original  $f(x)$ , e do PSF  $h(x)$ , respectivamente, e  $u$  é o vetor de coordenadas  $[u, v]^T$ . Separando as partes relacionadas a magnitude e fase, tem-se

$$\begin{aligned} |G(u)| &= |F(u)||H(u)| \text{ e} \\ \angle G(u) &= \angle F(u) + \angle H(u) \end{aligned} \quad (3.26)$$

Assumindo que o PSF  $h(x)$  do desfoque é centralizado simetricamente, isto é,  $h(x) = h(-x)$ , a sua transformada de Fourier será sempre real, e como consequência a sua fase será uma função bivalente dada por

$$\angle H(u) = \begin{cases} 0, & \text{se } H(u) \geq 0 \\ \pi, & \text{se } H(u) < 0 \end{cases} \quad (3.27)$$

isso significa que

$$\angle G(u) = \angle F(u) \text{ para todo } H(u) \geq 0 \quad (3.28)$$

Ou seja, a fase da imagem observada  $\angle G(u)$  nas frequências onde  $H(u)$  é positivo, é invariante a borramentos.

O LPQ usa a informação de fase local extraída a partir da transformada discreta 2D de Fourier ou, mais precisamente, a transformada de Fourier de curto termo (*Short-Term Fourier Transform* - STFT) computada ao longo de  $MXM$  vizinhos  $N_x$  em cada pixel  $x$  da imagem  $f(x)$  definido como

$$F(u, x) = \sum_{y \in N_x} f(x - y) e^{-j2\pi u^T y} = W_u^T f_x \quad (3.29)$$

onde  $W_u$  é o vetor base da DFT 2D na frequência  $u$ , e  $f_x$  é outro vetor contendo todos  $M^2$  amostragens da imagem a partir de  $N_x$ . A partir da Equação 3.29 é possível notar que uma maneira eficiente de implementar o STFT é usar convoluções 2D  $f(x) * e^{-2\pi j u^T x}$

para todo  $u$ . Uma vez que as funções base são separáveis, o processamento pode ser feito usando convoluções 1D para as linhas e colunas sucessivamente.

No LPQ somente quatro coeficientes complexos são considerados, correspondentes as frequências 2D  $u_1 = [a, 0]^T$ ,  $u_2 = [0, a]^T$ ,  $u_3 = [a, a]^T$  e  $u_4 = [a, -a]^T$  onde  $a$  é a frequência escalar abaixo da primeira passagem por zero de  $H(u)$  que satisfaz a Equação 3.29. Seja

$$\begin{aligned} F_x^c &= [F(u_1, x), F(u_2, x), F(u_3, x), F(u_4, x)], \text{ e} \\ F_x &= [Re\{F_x^c\}, Im\{F_x^c\}]^T \end{aligned} \quad (3.30)$$

onde  $Re\{\}$  e  $Im\{\}$  retorna as partes real e imaginária de um número complexo, respectivamente. A matriz de transformação  $8XM^2$  correspondente é definida como

$$W = [Re\{w_{u1}, w_{u2}, w_{u3}, w_{u4}\}, Im\{w_{u1}, w_{u2}, w_{u3}, w_{u4}\}]^T \quad (3.31)$$

portanto

$$F_x = W f_x \quad (3.32)$$

Ojansivu e Heikkilä assumem que a função da imagem  $f(x)$  é um resultado de um processo de Markov de primeira ordem para a análise estatística dos coeficientes, onde o coeficiente de correlação entre os valores do pixel adjacente é  $\rho$ , e a variância de cada amostra é  $\sigma^2$ . Sem perder a generalização é possível assumir que  $\sigma^2 = 1$ . Como resultado, a covariância entre as posições  $x_i$  e  $x_j$  são

$$\sigma_{ij} = \rho^{\|x_i - x_j\|} \quad (3.33)$$

onde  $\|\cdot\|$  denomina o modelo  $L_2$ , e a matriz de covariância de todas as amostras  $M$  em  $N_x$  pode ser expressa como

$$C = \begin{bmatrix} 1 & \sigma_{12} & \dots & \sigma_{1M} \\ \sigma_{21} & 1 & \dots & \sigma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1} & \sigma_{M2} & \dots & 1 \end{bmatrix} \quad (3.34)$$

Então, a matriz de covariância do vetor de coeficientes transformado  $F_x$  pode ser obtido a partir de

$$D = WCW^T \quad (3.35)$$

Sendo que  $D$  não é matriz diagonal para  $\rho > 0$ , significando que os coeficientes estão correlacionados.

Antes da quantização os coeficientes são descorrelacionados, devido que a informação é maximamente preservada na quantização escalar se as amostras a serem quantizadas são estatisticamente independentes. Assumindo uma distribuição gaussiana a independência é atingida utilizando a transformação *whitening*

$$G_x = V^T F_x \quad (3.36)$$

onde  $V$  é uma matriz ortogonal derivada do valor de decomposição singular (*Singular Value Decomposition* - SVD) da matriz  $D$

$$D = U \Sigma V^T \quad (3.37)$$

O  $G_x$  é processado para todas as posições da imagem e os vetores resultantes são quantizados usando um quantizador escalar

$$q_j = \begin{cases} 1, & \text{se } g_j \geq 0 \\ 0, & \text{caso contrário} \end{cases} \quad (3.38)$$

onde  $g_j$  é o componente  $j$ th de  $G_x$ . Os coeficientes quantizados são representados como valores inteiros entre 0-255 usando a codificação binária

$$b = \sum_{j=1}^8 q_j 2^{j-1} \quad (3.39)$$

Então, histograma desses valores inteiros de todas as posições da imagem é construído e usado como um vetor de características de 256 dimensões na classificação.

Descorrelação e quantização não causam nenhum efeito na propriedade invariante a desfoque. Na transformada *whitening* os coeficientes dos vetores são sujeitos a uma



rotação de oito dimensões que causa um deslocamento uniforme na fase de todos os vetores. Na quantização o espaço de oito dimensões é dividido em 256 hipercubos e a atribuição a um desses hipercubos depende somente das informações da fase [34].

## 3.3 Classificação e regressão

### 3.3.1 Máquina de vetores de suporte

A máquina de vetores de suporte (SVM) é uma técnica de aprendizagem de máquina que vem recebendo grande atenção nos últimos anos por possuir um desempenho considerável como classificador de grandes quantidades de dados e com capacidade de resolver problemas de classificação e regressão. Introduzida por Vladimir Vapnik e embasada na Teoria de Aprendizado Estatístico algumas de suas principais características que tornam seu uso atrativo são: (i) Boa capacidade de generalização medida por sua eficiência na classificação de dados que não pertençam ao conjunto utilizado no treinamento; (ii) Robustez em grandes dimensões, como imagens; (iii) Convexidade da função objetivo, por implicar na otimização de uma função quadrática que possui apenas um mínimo global que é uma vantagem quando comparado as Redes Neurais Artificiais em que há a presença de mínimos locais na função objetivo a ser minimizada; (iv) Teoria bem definida dentro da matemática e estatística.

É uma técnica de aprendizado estatístico baseada no princípio da Minimização do Risco Estrutural e pode ser usada para resolver problemas de classificação e regressão onde, por exemplo, dadas duas classes e um conjunto de pontos que pertencem a estas, as máquinas de vetores de suporte determinam o hiperplano que os separa de forma a colocar a maior quantidade possível de pontos da mesma classe do mesmo lado ao mesmo tempo em que a distância de cada classe a esse hiperplano é maximizada.

Normalmente, as máquinas de vetores de suporte operam em um espaço de dimensão maior que a dimensão dos dados originais no chamado espaço de características, onde se consegue a maximização desejada para obter uma melhor capacidade de generalização de classificação. Os resultados da aplicação desta técnica são comparáveis aos obtidos por outras técnicas de aprendizagem como as Redes Neurais Artificiais [35] e em algumas tarefas têm se mostrado superiores tal como na detecção de faces em imagens.

### 3.3.1.1 Definição matemática

Considerando os exemplos de treinamento  $\{(x_i, d_i)\}_{i=1}^N$ , onde  $x_i$  é o padrão de entrada para o  $i$ th exemplo e  $d_i$  corresponde as respostas desejadas, ou seja, a classe com os labels. Assume-se que o padrão ou classe representada pelo valor  $d_i = +1$  e a classe representada por  $d_i = -1$  são linearmente separáveis. A equação da superfície de decisão em forma de hiperplano que faz a separação dos dados é definida como

$$w^T x + b = 0 \quad (3.40)$$

onde  $x$  é o vetor de entrada,  $w$  é um vetor de pesos ajustáveis e  $b$  é a constante de bias. Pode-se então definir

$$\begin{aligned} w^T x + b &\geq 0 \text{ para } d_i = +1 \\ w^T x + b &< 0 \text{ para } d_i = -1 \end{aligned} \quad (3.41)$$

Para um dado vetor de pesos  $w$  e bias  $b$  a separação entre o hiperplano definido na Equação 3.40 e o ponto do dado mais próximo é chamado de margem de separação, definido como  $p$ . O objetivo de uma máquina de vetor de suporte é encontrar um hiperplano particular onde a margem de separação  $p$  é maximizada. Sob essas condições a superfície de decisão é dita como hiperplano ótimo. A Figura 20 ilustra a construção geométrica de um hiperplano ótimo para um espaço bidimensional.

Assumindo que  $w_0$  e  $b_0$  sejam os valores ótimos para o vetor de pesos e bias, respectivamente, o hiperplano ótimo representando a superfície de decisão linear multidimensional no espaço de entrada pode ser definido pela equação

$$w_0^T x + b_0 = 0 \quad (3.42)$$

a função discriminante

$$g(x) = w_0^T x + b_0 \quad (3.43)$$

dá uma medida algébrica da distância a partir de  $x$  até o hiperplano ótimo. Uma maneira

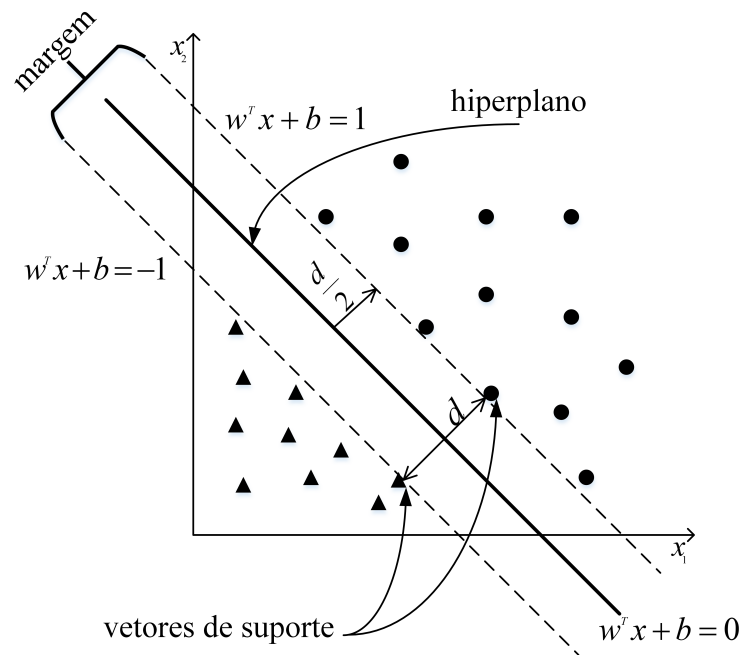


Figura 20: Exemplo de hiperplano ótimo para padrões linearmente separáveis.

fácil de visualizar isso é definir  $x$  como

$$x = x_p + r \frac{w_0}{\|w_0\|} \quad (3.44)$$

onde  $x_p$  é a projeção normal de  $x$  sobre o hiperplano ótimo, e  $r$  é a distância algébrica desejável;  $r$  é positivo se  $x$  está no lado positivo do hiperplano ótimo e negativo se  $x$  está no lado negativo. Sendo que, por definição,  $g(x_p) = 0$ , tem-se

$$g(x) = w_0^T x + b_0 = r \|w_0\| \text{ ou } r = \frac{g(x)}{\|w_0\|} \quad (3.45)$$

Particularmente, a distância a partir da origem (i.e.,  $x = 0$ ) até o hiperplano ótimo é dado por

$$\frac{b_0}{\|w_0\|} \quad (3.46)$$

Se  $b_0 > 0$ , a origem está no lado positivo do hiperplano ótimo; se  $b_0 < 0$ , está no lado negativo. Se  $b_0 = 0$ , o hiperplano ótimo passa através da origem. A interpretação geométrica é dada na Figura 21.

O problema é encontrar os parâmetros  $w_0$  e  $b_0$  para o hiperplano ótimo dado o conjunto

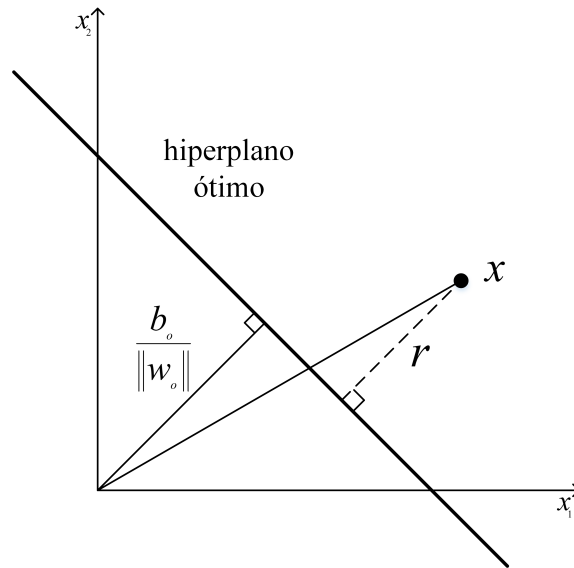


Figura 21: Interpretação geométrica da distância algébrica.

de treinamento  $\mathcal{T} = (x_i, d_i)$ . De acordo com a Figura 21 nota-se que o par  $(w_0, b_0)$  deve respeitar as seguintes condições:

$$\begin{aligned} w_0^T x_i + b_0 &\geq 1 \text{ para } d_i = +1 \\ w_0^T x_i + b_0 &\leq -1 \text{ para } d_i = -1 \end{aligned} \quad (3.47)$$

Os pontos particulares  $(x_i, d_i)$  onde a primeira ou segunda linha da Equação 3.47 é satisfeita são chamados de vetores de suporte. Os vetores de suporte são os dados que estão mais próximos da superfície de decisão e portanto são os mais difíceis de classificar. E, têm influência direta na localização ótima da superfície de decisão [35]. Considerando o vetor de suporte  $x^{(s)}$  onde  $d^{(s)} = +1$ . Por definição, tem-se

$$g(x^{(s)}) = w_0^T x^{(s)} \mp b_0 = \pm 1 \text{ para } d^{(s)} = \pm 1 \quad (3.48)$$

A partir da Equação 3.45 a distância algébrica do vetor de suporte  $x^{(s)}$  até o hiperplano ótimo é

$$r = \frac{g(x^{(s)})}{\|w_0\|} = \begin{cases} \frac{1}{\|w_0\|}, & \text{se } d^{(s)} = +1 \\ -\frac{1}{\|w_0\|}, & \text{se } d^{(s)} = -1 \end{cases} \quad (3.49)$$

onde o sinal positivo indica que  $x^{(s)}$  está no lado positivo do hiperplano ótimo e o sinal

de menos indica que  $x^{(s)}$  está no lado negativo. Considerando  $p$  o valor ótimo da margem de separação entre as duas classes que constituem o conjunto de treinamento  $\mathcal{T}$ . Então, a partir da Eq. Equação 3.49 tem-se

$$p = 2r = \frac{2}{\|w_0\|} \quad (3.50)$$

Onde afirma que maximizando a margem de separação entre as classes é equivalente a minimizar a norma Euclidiana do vetor de pesos  $w$ .

### 3.3.1.2 O truque do kernel

A versão básica do SVM permite apenas uma classificação linear, porém, isso pode ser alterado aplicando o truque do kernel (*kernel trick*). Os dados não-linearmente separáveis são transformados em um espaço com dimensão maior  $H$  utilizando uma função de mapeamento  $\varphi$  onde uma separação dos dados pode ser feita linearmente por um hiperplano, conforme pode ser observado na Figura 22. Devido  $H$  geralmente possuir um espaço de alta dimensionalidade o cálculo do produto escalar pode ser muito complexo. O truque do kernel é baseado nas observações onde  $\varphi$  aparece apenas nos produtos escalares e pode então ser trocado por uma função  $K$  que se comporta como um produto escalar e mantém o custo computacional baixo.  $K$  é chamado de função kernel e é definido como  $K(x_i, x_j) = [\varphi(x_i), \varphi(x_j)]$ . Algumas funções kernel básicas utilizadas são definidas na Tabela 3.

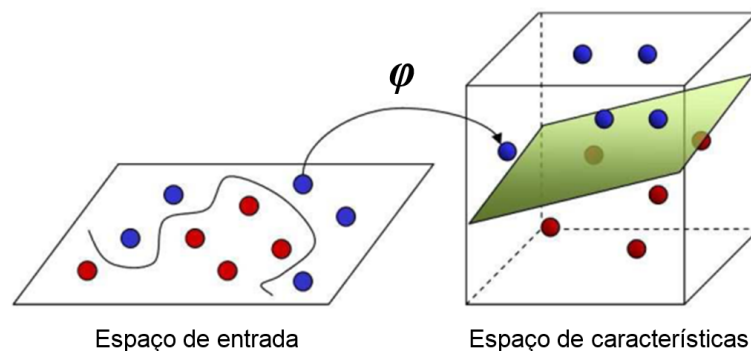


Figura 22: Mapeamento não-linear  $\varphi(\cdot)$  do espaço de entrada (2D) para o espaço de características (3D).

Tabela 3: Funções kernel básicas

Kernel	Produto escalar
Linear	$K(x_i, x_j) = (x_i \cdot x_j)$
Gaussiano	$K(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ ^2}{2\sigma^2}\right)$
Polinomial	$K(x_i, x_j) = (x_i^T x_j + 1)^d$
Tangente hiperbólica	$K(x_i, x_j) = \tanh(x_i \cdot x_j - \theta)$

### 3.3.1.3 Margens suaves

A extensão margem suave (*soft margin*) foi introduzida por C. Cortes e V. Vapnik para tratarem problemas causados por dados discrepantes (*outliers*) e ruídos que podem fazer com que o SVM aprenda um hiperplano com uma margem pequena entre as classes e assim prejudicar o objetivo de construir um modelo generalizado. Para prevenir isso, o problema básico de otimização foi estendido para

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \text{ sujeito a } c_i[(w, x_i) + b] \geq 1 - \xi_i, \forall i, \xi_i \geq 0 \quad (3.51)$$

onde  $\xi_i$  mede os erros de classificação das amostras  $x_i$  e é chamada de variável de folga (*slack variable*). Portanto, essa nova versão do SVM busca pelo hiperplano ótimo para separar os dados o melhor possível mas tolera erros de classificação para maximizar a margem entre as classes, conforme ilustrado na Figura 23. O peso entre essas duas medidas é controlado pelo parâmetro custo  $C$ .

### 3.3.1.4 Multiclasses

A limitação de construir SVM com mais de duas classes é normalmente resolvido pela divisão do problema em múltiplos problemas binários. Existem dois métodos comuns para a construção de classificadores multiclasses: um-contra-todos (*one-versus-all*) e um-contra-um (*one-versus-one*).

A estratégia um-contra-todos consiste na criação de  $N$  SVMs, sendo  $N$  o número total de classes do problema. Para o treinamento de cada um dos SVMs a classe respectiva a esse SVM é fixada e todas as demais são consideradas como a outra classe do problema. Ao final, para a classificação de uma nova instância este deve passar por todos os SVMs e, finalmente, o SVM que apresentar o valor máximo de classificação será definido como o valor de saída. Um ponto fraco do método um-contra-todos é que este utiliza

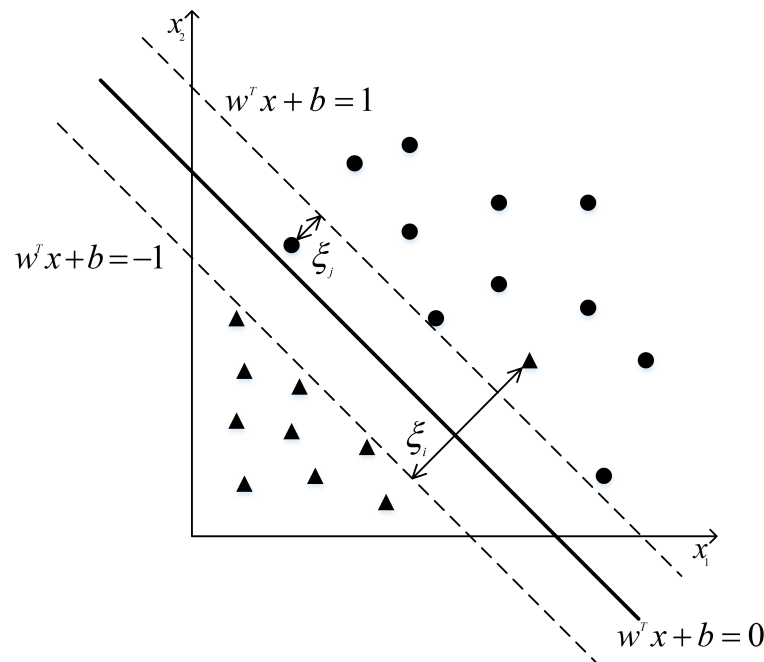


Figura 23: Exemplo de variáveis de folga em um SVM com margens suaves.

toda a população de treinamento em cada um dos SVMs. Desta forma, a velocidade de treinamento dos SVMs é baixa e o custo computacional elevado.

O método um-contra-um faz uso de SVMs de classificação binária. Cada um desses SVMs faz a classificação entre duas das classes presentes no problema. Para a classificação de um novo padrão o mesmo passa por todos os SVMs e em cada um deles receberá um voto para classificá-lo a uma das duas classes. Ao final, ele será considerado como pertencente da classe que recebeu mais votos durante o processo. A maior fraqueza do método é o número de SVMs. Os SVMs utilizam, neste caso, apenas um subconjunto dos dados de treinamento para definição dos vetores de suporte o que acelera o processo de treinamento. Em compensação, sabendo que para cada SVM existem parâmetros para serem definidos, a quantidade elevada de SVMs pode tornar esse processo altamente complexo.

### 3.3.2 Máquina de vetores de suporte para regressão

As máquinas de vetores de suporte para regressão (SVR) foram introduzidas por Vapnik brevemente após a extensão das margens suaves para SVM. O objetivo é encontrar um hiperplano mais plano. Para esse propósito, todos os dados de treinamento com um erro menor que  $\epsilon$  são ignorados (Figura 24). Portanto, desvios das amostras próximos ao hiperplano não são incluídos na função de perda (*loss function*). O conjunto de treinamento

é conseqüentemente reduzido levando a um novo problema de otimização:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \text{ sujeito a } \begin{cases} c_i - (w, x_i) - b \leq \epsilon \\ (w, x_i) + b - c_i \leq \epsilon \\ \forall i \geq 0 \end{cases} \quad (3.52)$$

Para combinar o SVR com a extensão de margens suaves duas variáveis de folga  $\xi_i$  e  $\xi_i^*$  são introduzidas, conforme ilustradas na Figura 24). O problema de otimização modificado a ser resolvido é:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \text{ sujeito a } \begin{cases} c_i - (w, x_i) - b \leq \epsilon + \xi_i \\ (w, x_i) + b - c_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \\ \forall i \geq 0 \end{cases} \quad (3.53)$$

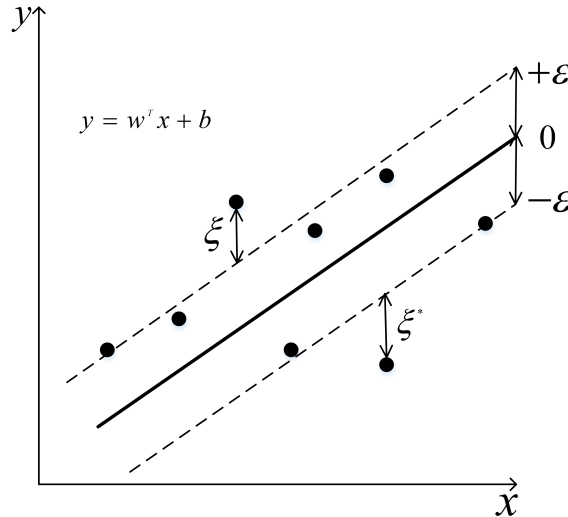


Figura 24: Exemplo de SVR linear com a margem de erro  $\epsilon$  e variáveis de folga.

Este capítulo apresentou, além de um breve estudo sobre o envelhecimento facial, uma revisão de algumas das principais técnicas para a extração de características de textura que descrevem rugas faciais, manchas na pele e formato facial. Também, as técnicas máquinas de vetores de suporte e máquinas de vetores de suporte para regressão foram apresentadas com o intuito de melhorar o entendimento da teoria para a posterior aplicação na abordagem da estimativa da idade através da combinação das duas técnicas.



## 4 Método proposto

A abordagem proposta para a estimativa da idade a partir da análise de imagens faciais é composta de extração de características globais (forma e aparência) e locais (rugas e manchas na pele). As características globais são extraídas utilizando AAM e as locais utilizando LBP, LPQ e Gabor *wavelets*. Várias combinações entre características globais e locais assim como alterações em seus parâmetros são realizadas com o intuito de encontrar a melhor representação do envelhecimento facial. Também, a estimativa hierárquica proposta é composta primeiramente pela classificação de uma imagem facial em faixas etárias utilizando SVM para então fazer a estimativa específica da idade utilizando regressão com o SVR para a definição de funções relacionadas à idade. O erro gerado na classificação em faixas etárias causado pelos limites rígidos do classificador é compensado na estimativa específica da idade pela sobreposição flexível de faixas etárias na etapa de regressão.

O método proposto é ilustrado no diagrama da Figura 25. Dada uma imagem colorida da face de um indivíduo o processo de estimação da idade inicia com a conversão desta imagem para níveis de cinza e é realizada uma transformação de similaridades para normalizar a imagem. As características locais são extraídas a partir das  $N$  regiões de interesse cortadas da imagem normalizada para a análise local da textura facial. Enquanto as características globais são extraídas utilizando os 68 pontos de referência para gerar o modelo de aparência ativa. Em seguida, as características globais e locais são combinadas para uma melhor representação do envelhecimento facial e posteriormente classificadas dentro de uma das  $N$  faixas etárias (classes) definidas. Então, de acordo com a faixa etária que o indivíduo for classificado é selecionado o regressor SVR específico treinado com regiões sobrepostas para compensar eventuais erros causados durante a classificação. Finalmente, a idade aparente através de um número inteiro é estimada.

Nas seções seguintes serão descritas detalhadamente as etapas para a estimação etária como a definição dos conjuntos de dados, o pré-processamento das imagens faciais, a extração das características globais e locais, a fusão e redução da dimensão das

características, e finalmente a classificação hierárquica para a estimativa da idade.

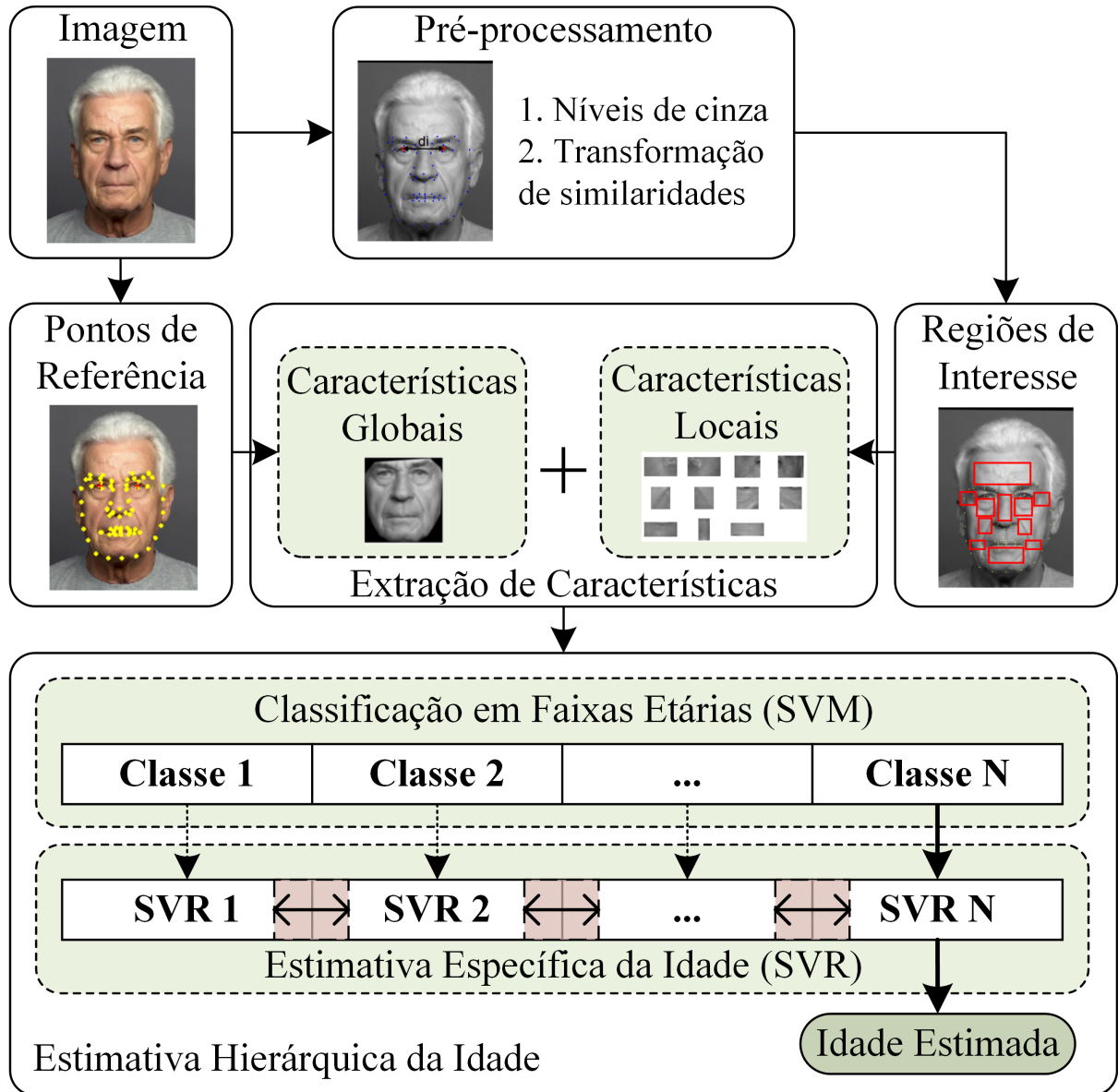


Figura 25: Método proposto para a estimativa da idade baseada na análise de imagens faciais.

## 4.1 Conjuntos de dados

A Tabela 2 apresentada no final do Capítulo 3.3.1 mostra que grande parte das pesquisas de estimação etária utilizaram as bases de dados FG-NET *Aging* e a base oriental YGA. As bases de dados escolhidas para esta pesquisa foram a FG-NET *Aging*, mas a distribuição de idades da base é significativamente tendenciosa para crianças [1], e a base MORPH *Album 2* a qual contém imagens faciais de indivíduos de 15 a 77 anos e que não foi muito

explorada para a estimaco etria.

#### 4.1.1 FG-NET *Aging*

A base de dados pblica FG-NET *Aging* (*Face and Gesture Recognition Research Network*) [36] contm 1.002 imagens faciais de 82 pessoas (homens e mulheres) de mltiplas etnias com uma mdia de 12 imagens por pessoas com idades variando de 0 a 69 anos, porm 50% das pessoas esto na faixa etria de 0 a 13 anos. As imagens so coloridas ou em nveis de cinza tendo uma dimenso mdia de  $384 \times 487$  pixels e resoluo variando entre 200 ppp a 1200 ppp com variaes de iluminao, pose, expresso facial e ocluses (e.g., bigode, barba, culos etc.). A FG-NET disponibiliza um arquivo contendo 68 pontos de referncia (*landmarks*) para cada imagem facial os quais foram anotados manualmente com preciso e tambm as idades cronolgicas de cada indivduo. Os 68 pontos de referncia podem ser visualizados na Figura 27 e suas descries na Tabela 4 que mostra os pontos de referncia de acordo com os elementos faciais. A Figura 26 mostra alguns exemplos de imagens faciais que compe a base de dados FG-NET. A distribuio das amostras por idade da FG-NET *Aging*  mostrada na Figura 28.



Figura 26: Exemplos de imagens faciais da base de dados FG-NET.

#### 4.1.2 MORPH *Album2*

Outra base de domnio pblico utilizada conforme mostra a Tabela 2  a MORPH *Album 2* [37] a qual contm 55.608 imagens faciais de 13.000 pessoas (homens e mulheres) variando de 15 a 77 anos. A base de dados foi coletada pelo *Face Aging Group* da Universidade da Carolina do Norte para a finalidade de aplicaes relacionadas a biometria facial. As imagens so coloridas e possuem fundo uniforme na cor cinza. A dimenso mdia  de  $400 \times 480$  pixels e resoluo variando entre 81 ppp a 96 ppp com variao de iluminao, pose, expresso facial e ocluses (e.g., bigode, barba, culos etc.). Grande parte das imagens so de afro-americanos, porm contm imagens de caucasianos e de outras etnias tambm.

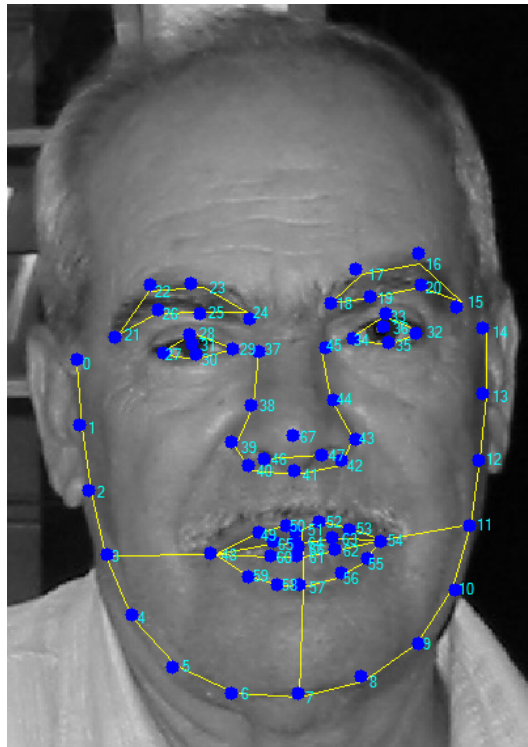


Figura 27: Face com os 68 pontos de referência (*landmarks*).

Tabela 4: Pontos de referência facial (*landmarks*) de acordo com a Figura 27.

Elementos faciais	Pontos de referência	Número de pontos
Contorno da face	0-14	15
Sobrancelha direita	15-20	6
Sobrancelha esquerda	21-26	6
Olho esquerdo	27-30	4
Pupila esquerda	31	1
Olho direito	32-35	4
Pupila direita	36	1
Nariz	37-45	9
Centro do nariz	67	1
Narinas	46-47	2
Contorno da boca	48-59	12
Lábio inferior	48, 60, 61, 62, 54	5
Lábio superior	54, 63, 64, 65, 48	5

A MORPH disponibiliza um arquivo com metadados contendo informações sobre a idade, gênero, etnia, altura e peso.

Chang et al. selecionaram 5.492 imagens de pessoas caucasianas a partir da base de dados MORPH *Album 2* para sua pesquisa sobre estimação etária e anotaram cada uma das imagens com os mesmos 68 pontos de referência ilustrados na Figura 27 de

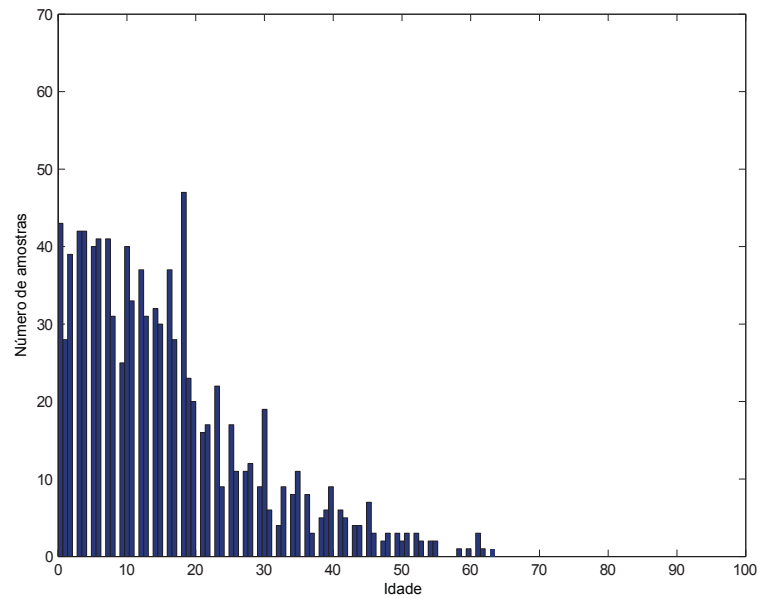


Figura 28: Distribuição das amostras por idade (FG-NET).

forma automática [38]. Pelo fato dos autores disponibilizarem publicamente as imagens e os pontos de referência, foi decidido utilizar o mesmo conjunto de dados neste trabalho. A Figura 29 mostra alguns exemplos de imagens faciais que compõe a base de dados MORPH. A distribuição das amostras por idade da MORPH *Album 2* para as imagens selecionadas é mostrada na Figura 30.



Figura 29: Exemplos de imagens faciais da base de dados MORPH.

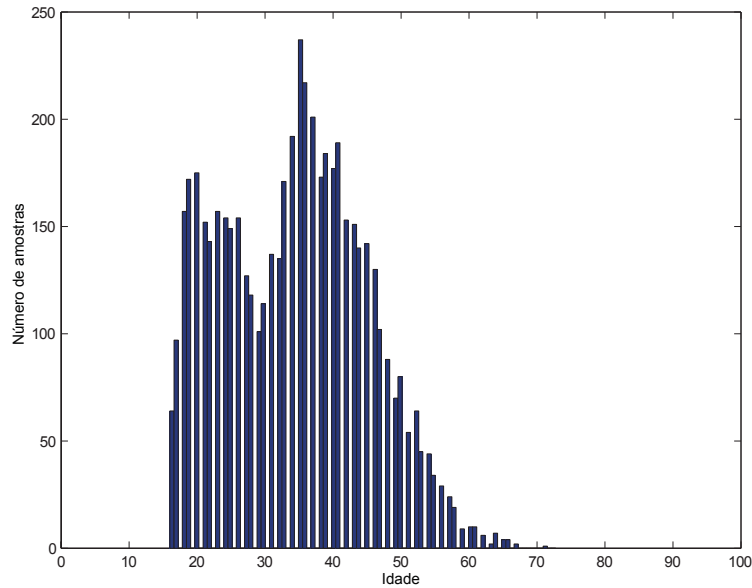


Figura 30: Distribuição das amostras por idade do conjunto de dados selecionado (MORPH).

## 4.2 Pré-processamento

O pré-processamento é apenas realizado antes da extração das características locais, devido o método AAM já integrar os pré-processamentos necessários, como transformação de similaridades para a correção da pose e escala, e normalização da iluminação para a extração das características globais, conforme descritos na subseção 3.2.1. As imagens da base de dados FG-NET *Aging* foram capturadas utilizando diferentes métodos como digitalização de fotos e fotos tiradas a partir de câmeras digitais, sendo algumas coloridas e outras na escala de cinza. Portanto, todas as imagens foram convertidas para níveis de cinza primeiramente, uma vez que a informação da cor não é relevante para essa pesquisa e também devido ao custo computacional ser reduzido.

Para cada imagem dos conjuntos selecionados FG-NET *Aging* e MORPH *Album 2* são fornecidos 68 pontos de referência (*landmarks*) incluindo a posição do centro dos olhos, conforme ilustrados na Figura 31. De modo a melhorar a precisão da localização das áreas de interesse para a extração das características locais a técnica de transformação de similaridades (*nonreflective similarity transformations*) foi utilizada para a normalização de todas as imagens faciais baseada na posição dos olhos, de acordo com o trabalho de estimação etária proposto por Han et al. [1]. Na transformação de similaridades uma simples expressão matemática é aplicada em toda a imagem, ou seja, uma transformação global, onde se deseja corrigir as distorções causadas pela rotação e escala da imagem

facial em relação a um referencial. Para isso, dispõe-se de uma imagem facial inicial em um sistema de pixels (discreto) e se quer chegar a outra imagem facial digital, porém retificada com um sistema de coordenadas também discreto. Para executar o ajustamento da imagem são necessários alguns pontos de controle (no caso foi utilizado o centro dos dois olhos) onde deve-se conhecer suas coordenadas no sistema da imagem retificada. Através de um ajustamento pelo método paramétrico são calculados os coeficientes de transformação entre ambos os sistemas que inclui rotação, escala e translação da imagem preservando formas e ângulos [10] para finalmente ter uma imagem facial normalizada com a mesma distância interpupilar, conforme ilustrada na Figura 32. As imagens da base de dados MORPH *Album 2* são provenientes de *mugshots* (fotos de identificação), portanto, assim como a FG-NET, não foram tiradas em ambientes bem controlados sendo necessário o mesmo pré-processamento descrito anteriormente, para principalmente, corrigir a pose e a escala da imagens. Com as imagens normalizadas é possível aplicar uma mesma máscara em todas as imagens faciais para delimitar e recortar as regiões de interesse que serão descritas na próxima seção.

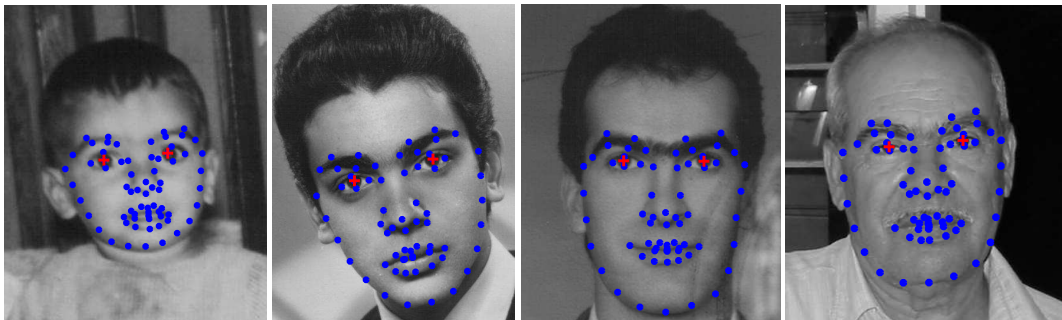


Figura 31: Imagens em níveis de cinza com os 68 *landmarks* (2, 17, 28 e 69 anos respectivamente) (FG-NET).

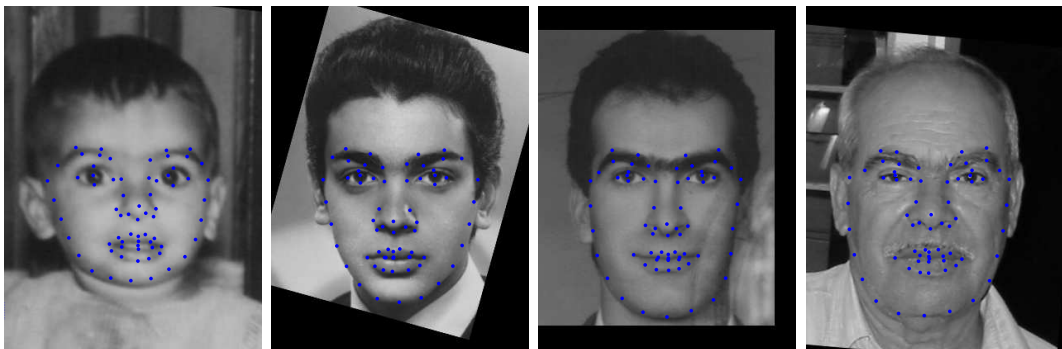


Figura 32: Imagens redimensionadas e rotacionadas utilizando transformação de similaridades (2, 17, 28 e 69 anos respectivamente) (FG-NET).

O método LBP utiliza padrões binários uniformes para lidar com possíveis ruídos



nas imagens, conforme descrito na seção 3.2.3, assim como o método LPQ também é robusto a ruídos na imagem. Portanto, os ruídos impulsivos (sal e pimenta), gerados no processo de rotação da imagem durante a transformação de similaridades foram desconsiderados neste trabalho. Também, não foram aplicadas equalizações para balancear o contraste da imagem, pois os métodos LBP e LPQ são robustos a variações de iluminação.

### 4.3 Regiões de interesse

Para a extração das características locais é necessário a localização de componentes faciais correspondentes às regiões que podem conter rugas, como o contorno dos olhos e boca, bochechas, queixo, nariz e testa, além de marcas que aparecem conforme a pessoa vai ficando idosa como sardas, manchas na pele pela alta exposição à raios ultravioleta do sol, rugas muito finas, entre outras definidas de acordo com a classificação de rugas faciais proposta por Lemperle [12] e ilustrada na Figura 33.

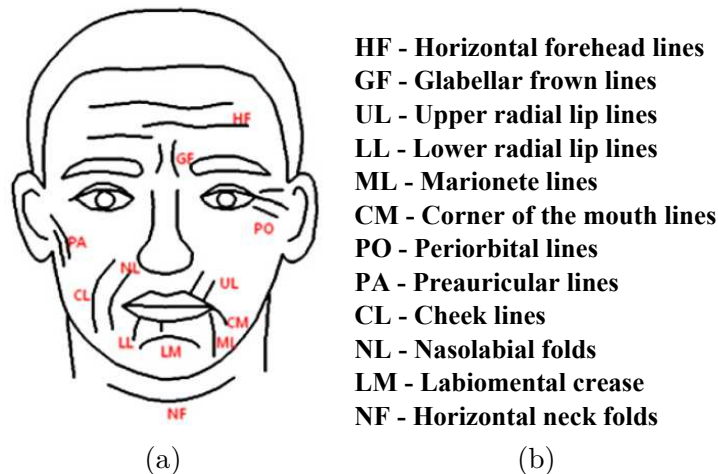


Figura 33: Classificação de rugas faciais proposta por Lemperle [12]: (a) posição das rugas e (b) descrição.

As seguintes regiões de interesse foram escolhidas para a análise local da textura: cantos da boca, cantos dos olhos (linhas periorbital), cantos do nariz (sulcos nasolabial), bochechas, queixo, topo do nariz e testa, totalizando em 11 regiões de interesse que cobrem áreas faciais contendo as rugas definidas na Figura 33a. Foram utilizados alguns pontos de referência (*landmarks*) para definir os retângulos contendo as áreas de interesse. Por exemplo, para a definição do retângulo da região da testa foram utilizados dois *landmarks* no topo das duas sobrancelhas como referência para a largura do retângulo enquanto a altura foi definida de forma a cobrir as linhas horizontais da testa. As 11 regiões de interesse para a extração de características locais são ilustradas na Figura 34.



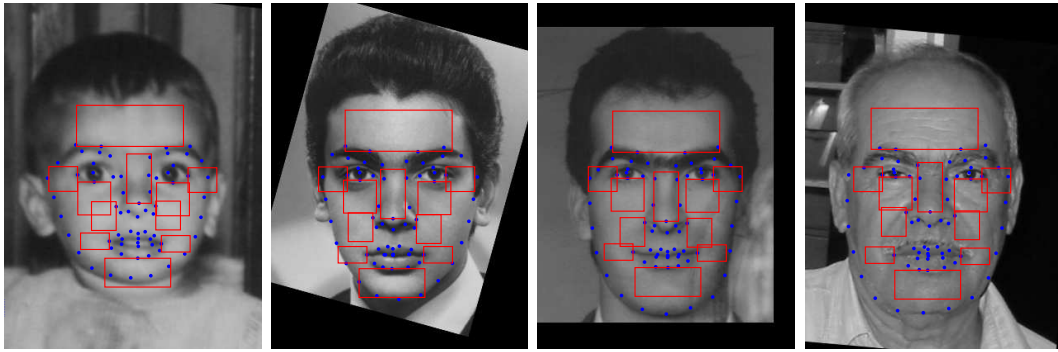


Figura 34: As 11 regiões de interesse para a extração de características locais (2, 17, 28 e 69 anos respectivamente) (FG-NET).

O recorte das áreas de interesse é essencial à normalização, pois elimina inúmeros artefatos desnecessários e preserva as informações e características mais relevantes aos estudos sobre estimacão etária deste trabalho. Portanto, as 11 regiões de interesse são cortadas para serem aplicados os extratores de características locais. Devido as imagens serem normalizadas, conforme descrito na seção 4.2, todos os retângulos cortados das áreas de interesse possuem as mesmas resoluções, como por exemplo, a região da testa possui uma resolução de  $131 \times 51$  para todas as imagens dos conjuntos de dados, conforme mostram as Figuras 35 a 38, as quais mostram as regiões de interesse cortadas de uma criança de 2 anos, adolescente de 17 anos, homem de 28 anos e de um idoso de 69 anos de idade, respectivamente.

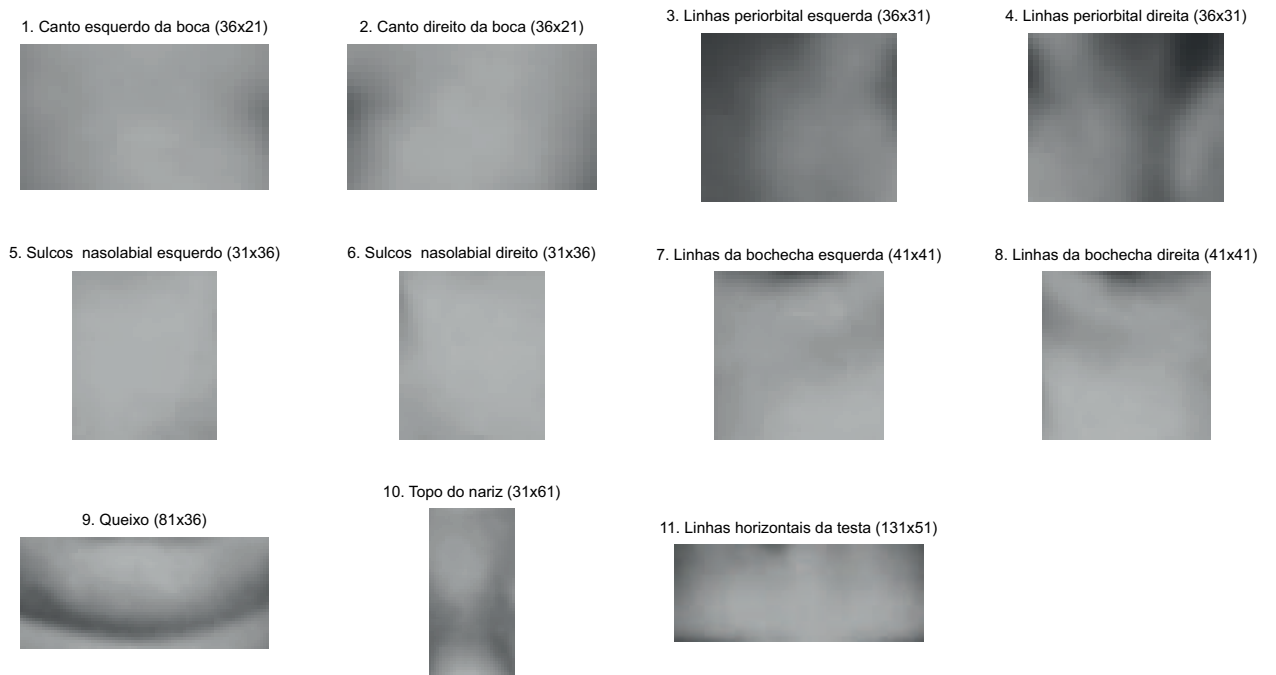


Figura 35: Regiões de interesse (2 anos) (FG-NET).

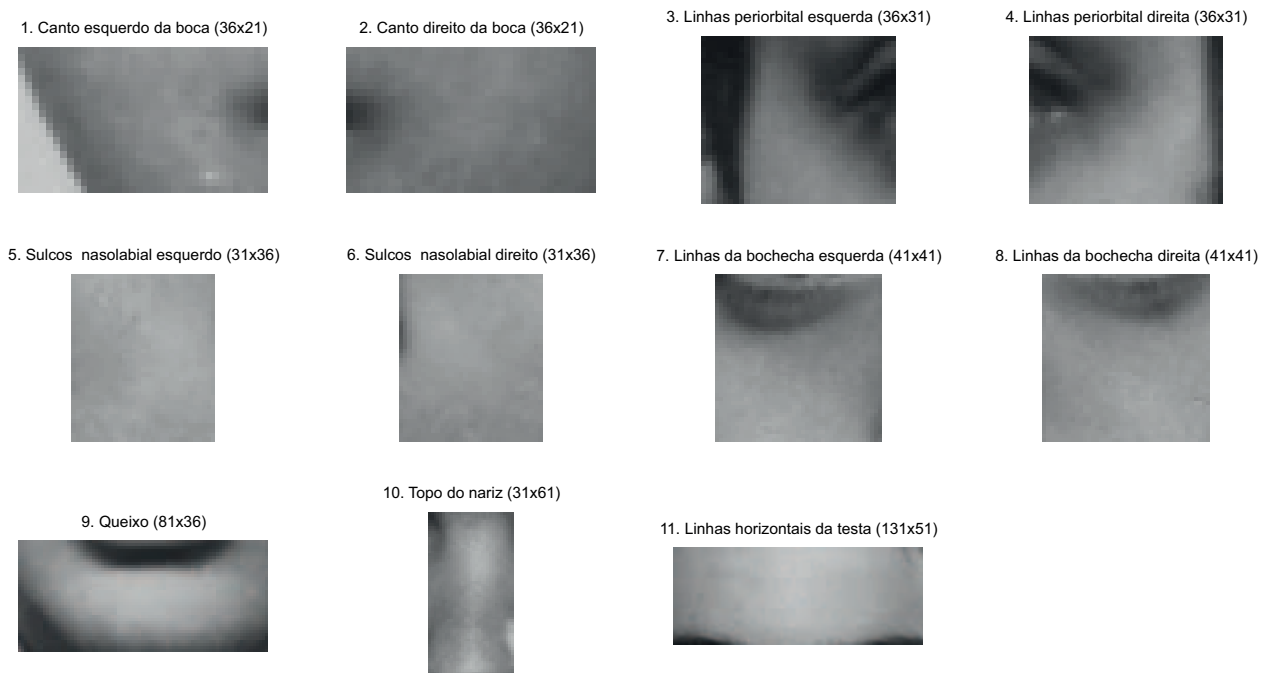


Figura 36: Regiões de interesse (17 anos) (FG-NET).

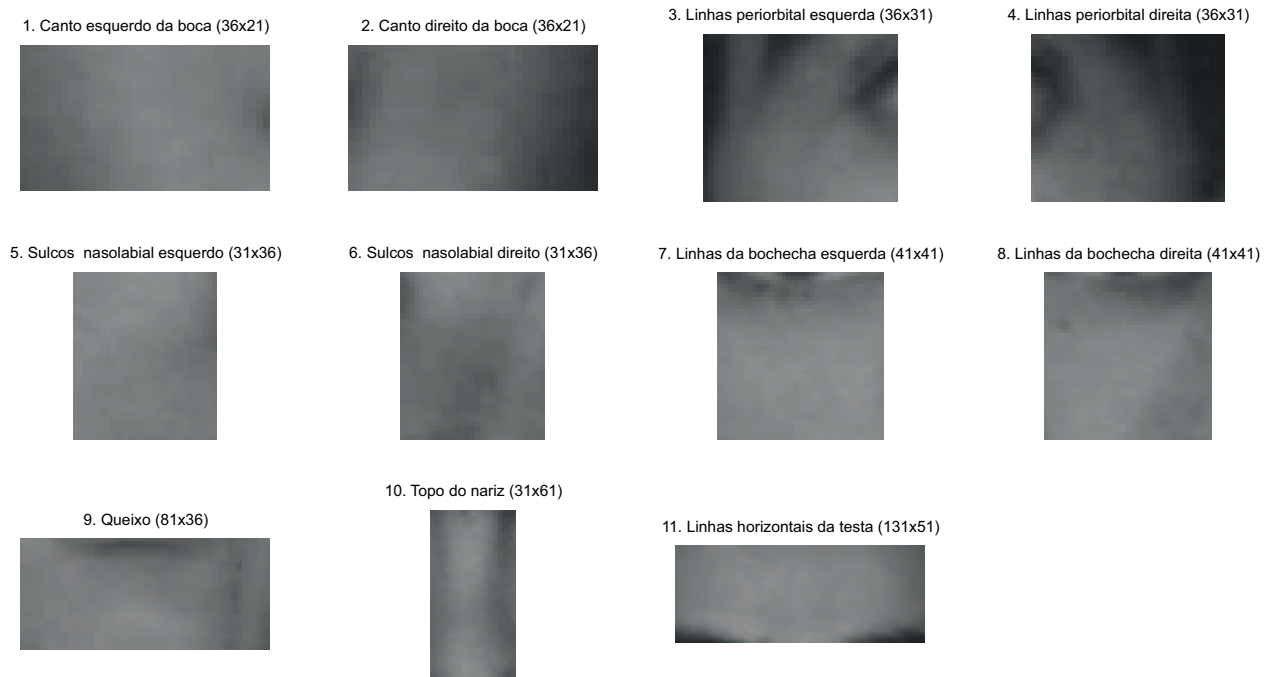


Figura 37: Regiões de interesse (28 anos) (FG-NET).

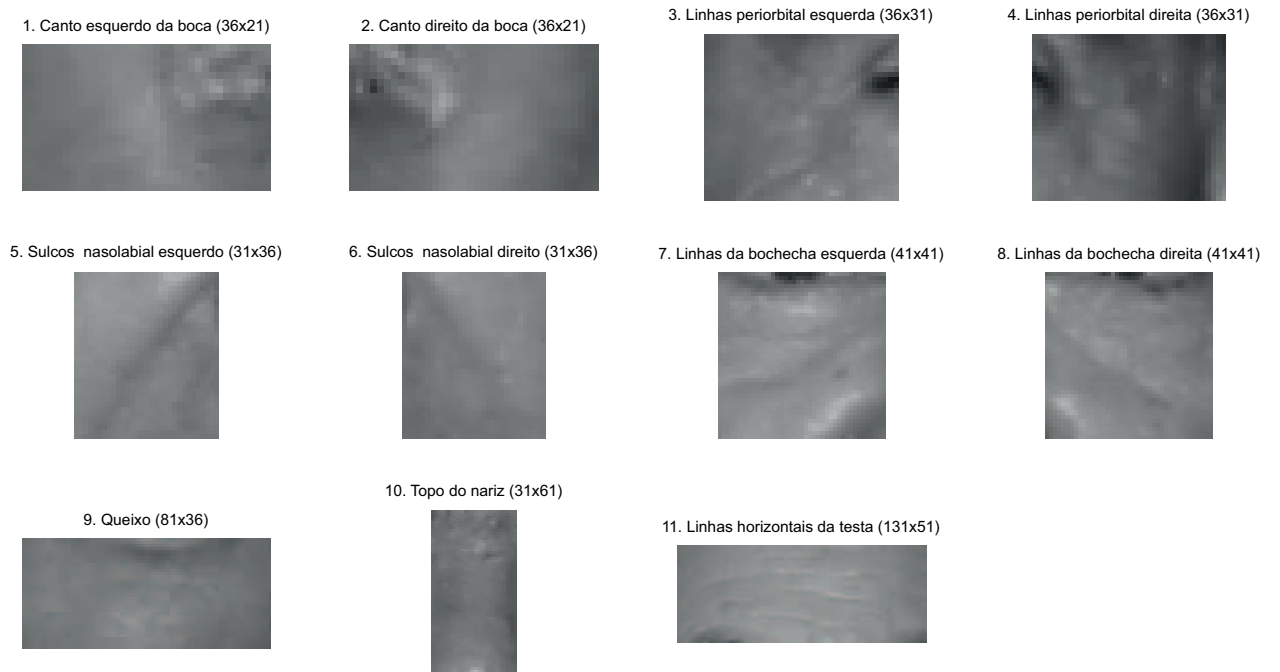


Figura 38: Regiões de interesse (69 anos) (FG-NET).

## 4.4 Extração de características

### 4.4.1 Características globais

As características globais utilizadas são derivadas dos parâmetros combinados do modelo de aparência ativa (AAM) que é um modelo paramétrico que contém o formato e aparência da face humana o qual é modelado utilizando análise de componentes principais (PCA), sendo largamente utilizado para o modelamento da face e extração de características [12]. Para a construção do AAM são utilizados os 68 pontos de referência disponibilizados pelos conjuntos FG-NET e MORPH, e não é necessário realizar pré-processamentos, pois o método já aplica transformação de similaridades para alinhar as imagens e também é feita uma normalização da iluminação, conforme descritos na seção 3.2.1.

### 4.4.2 Características locais

As características locais relacionadas às rugas são extraídas utilizando um conjunto de filtros de Gabor por serem robustos a ruídos como cabelos, barbas, bigodes e sombras como descrito na seção 3.2.2. As características locais relacionadas a detalhes finos da pele (rugos finas, marcas etc.) são extraídas utilizando LBP por ser um simples e eficiente extrator de características baseado em operadores de textura o qual rotula regiões de uma imagem limiarizando os vizinhos de cada pixel e considera o resultado como um número binário. A propriedade mais importante do operador LBP em aplicações reais é a robustez a mudanças na iluminação e a ruídos, pois considera padrões binários uniformes, conforme descrito na seção 3.2.3. O operador LPQ o qual é similar ao LBP, porém robusto a imagens com borramentos também é utilizado, como descrito na seção 3.2.4, para capturar rugas e demais marcas na pele através de características de fase local. Todas essas técnicas para a descrição da textura da pele são aplicadas nas 11 regiões de interesse recortadas e normalizadas pelo processo de pré-processamento descrito anteriormente. Após a extração das características de cada uma das 11 regiões, os vetores de características resultantes são concatenados para gerar um vetor contendo as descrições faciais das 11 regiões locais para representar a imagem facial.

## 4.5 Fusão das características

Para combinar múltiplas características normalmente derivadas de técnicas diferentes e com unidades diferentes as técnicas de combinação por nível de características, nível de

escores e nível de decisões têm sido utilizadas com sucesso em aplicações na área de biometria [39]. A combinação por nível de características contém informações relevantes sobre os dados brutos. Também, a técnica é capaz de eliminar informações redundantes a partir da correlação das características concatenadas [12]. Portanto, é utilizada nesta pesquisa para combinar características globais e locais.

Dada uma imagem facial as características globais através dos parâmetros combinados do AAM ( $f_{AAM} = (c_1, \dots, c_k)$ ) são extraídas, assim como as características locais através de GW ( $f_{GW} = (w_1, \dots, w_l)$ ), LBP ( $f_{LBP} = (t1_1, \dots, t1_m)$ ) e LPQ ( $f_{LPQ} = (t2_1, \dots, t2_n)$ ), sendo  $k$ ,  $l$ ,  $m$  e  $n$  as dimensões originais dos vetores. Devido as características locais possuírem uma alta dimensionalidade em seus vetores originais, apenas os componentes principais são selecionados através do método PCA para que suas dimensões sejam reduzidas. Para a normalização dos vetores de características o método z-escore (*z-score*) é utilizado [39]. Em estatística o z-escore (ou escore padrão) é usado para comparar médias de conjuntos de dados diferentes homogeneamente distribuídos. O escore indica quantos desvios padrões uma observação está acima ou abaixo da média, sendo útil em pesquisas utilizando análise estatística porque permite a comparação de valores de observações de diferentes distribuições normais. De fato, quando instâncias de diferentes conjuntos de dados são transformados em z-escores eles tornam-se passíveis de serem comparados. Para o cálculo do z-escore são necessários a média e o desvio padrão de acordo com a Equação 4.1 definida abaixo

$$f'_i = \frac{f_i - \mu}{\sigma_i}, \quad i = AAM, GW, LBP, LPQ \quad (4.1)$$

onde  $f_i$  é o vetor de características (AAM, GW, LBP ou LPQ),  $\mu_i$  e  $\sigma_i$  são a média e o desvio padrão do vetor de característica selecionado, respectivamente,  $f'_i$  é vetor de características normalizado. A fusão das características é feita pela concatenação dos vetores normalizados como segue

$$f_{fused} = (f'_{AAM} f'_{GW} f'_{LBP} f'_{LPQ}) = (c_1, \dots, c_k, w_1, \dots, w_l, t1_1, \dots, t1_m, t2_1, \dots, t2_n) \quad (4.2)$$

onde  $f_{fused}$  é o vetor com as características combinadas a serem utilizadas na etapa de redução do espaço de características onde será descrito na próxima seção. Durante a combinação, não são aplicadas ponderações nos conjuntos de características para não deixá-los dependentes dos conjuntos de dados, ou seja, muito específico, pois o peso das

características extraídas pelo AAM para o conjunto de dados FG-NET pode não ter o mesmo o peso para o conjunto MORPH.

## 4.6 Redução do espaço de características

Geralmente, algoritmos de redução da dimensão do espaço de características são utilizados após a extração utilizando Gabor *wavelets* onde os resultados da convolução de todos os filtros são concatenados fazendo com que a dimensão do espaço de características seja elevada, portanto, tais técnicas são aplicadas para aumentar a eficiência computacional reduzindo a dimensão das características.

A maioria dos estudos anteriores tem utilizado a técnica PCA para a redução das dimensões das características extraídas com Gabor *wavelets*. Muitos pesquisadores têm selecionado apenas características que sejam necessárias ao invés de utilizar todas. Portanto, métodos de redução são mais convenientes para a seleção de características.

A técnica PCA é utilizada nessa pesquisa para a redução da dimensionalidade dos vetores de características derivados dos filtros de Gabor, LBP e LPQ e também devido a concatenação de características locais e globais que aumentam a dimensão do espaço de características e que levam a um aumento no custo computacional e memória do sistema. Portanto, a dimensão das características concatenadas, de acordo com o método descrito anteriormente, é reduzida utilizando PCA conforme definido abaixo

$$f_{PCA} = (f_1, \dots, f_d), \quad d < k + l + m + n \quad (4.3)$$

onde  $d$  é a nova dimensão das características concatenadas e  $f_{PCA}$  é o vetor de características a ser utilizado na etapa da estimativa da idade. A Figura 39 mostra o diagrama do método para a fusão de características globais e locais.

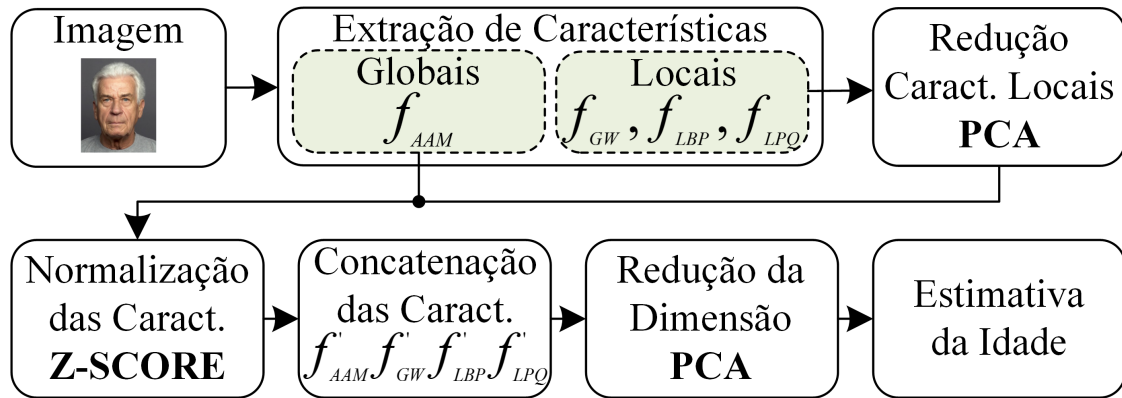


Figura 39: Método para a fusão de características globais e locais.

## 4.7 Estimativa da idade

A distribuição de características relacionadas à idade difere de acordo com a faixa etária, rugas, por exemplo, são normalmente encontradas na fase adulta, já características geométricas mudam durante a fase de crescimento. Por causa disso, as características faciais que são específicas de uma determinada faixa etária podem não ser consideradas quando treinadas através de um estimador *single-level*, o qual é treinado a partir de uma faixa única de idades, podendo assim afetar o desempenho da estimativa etária. Portanto, os classificadores hierárquicos estão sendo utilizados para melhorar o desempenho na estimativa da idade [12]. Porém, o método de estimativa hierárquica usando classificadores para faixas etárias que não incluem sobreposição da idade entre as faixas possui um inconveniente. Caso uma imagem facial de um indivíduo com 15 anos de idade, por exemplo, seja classificada dentro de uma faixa etária de 20-39 anos durante a etapa de classificação em faixas que não incluem sobreposição de faixas o erro da estimativa será provavelmente maior do que o erro de um classificador *single-level* e o erro será propagado para a etapa da estimativa específica da idade utilizando regressão.

O método proposto consiste em realizar primeiramente a classificação de uma imagem facial em faixas etárias utilizando SVM para em seguida fazer a estimativa específica da idade utilizando regressão com o SVR o qual define as funções relacionadas à idade (Figura 25). Também, são utilizadas faixas sobrepostas entre duas faixas adjacentes na etapa de regressão, similares aos classificadores fuzzy, para reduzir os erros gerados na etapa de classificação, como ilustrado na Figura 40. As faixas sobrepostas são determinadas pelos erros de idade gerados durante a classificação e podem ser flexíveis para cada SVR.

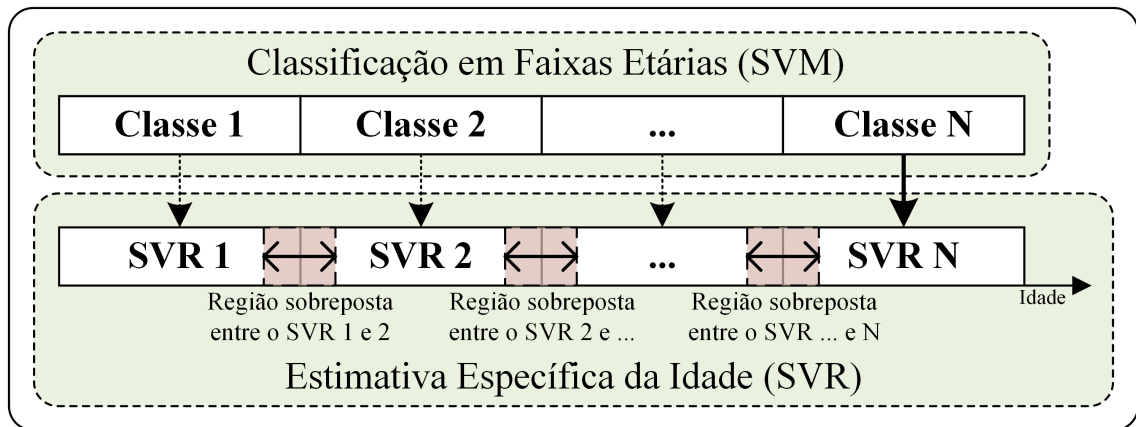


Figura 40: Método de classificação hierárquica com regiões sobrepostas na etapa de estimativa específica da idade.

Este capítulo descreveu detalhes do método proposto para a construção de um classificador hierárquico para a estimativa etária baseado em características globais e locais extraídas a partir de imagens faciais. Inicialmente, uma visão geral do método foi apresentado seguido da definição dos conjuntos de dados utilizados, pré-processamento das imagens, definição das áreas de interesse para a extração das características locais, as técnicas a serem utilizadas para a extração das características globais e locais, assim como suas combinações e redução da dimensionalidade, para finalmente descrever o método de estimativa da idade. O próximo capítulo descreve os resultados obtidos a partir do método aqui proposto.



## 5 Resultados experimentais

Uma vez que os experimentos conduzidos nessa pesquisa foram estritamente computacionais, os materiais e métodos utilizados para obter os resultados desse trabalho podem ser descritos em termos de configuração de hardware e software. Os algoritmos para extração de características, redução da dimensionalidade do espaço de características, classificação e regressão foram implementados no ambiente de programação MATLAB versão R2013b e todos os experimentos foram rodados no mesmo computador o qual possui um processador Intel® Core™ i5-3337U (3 MB Cache, 1.80 GHz, 4 cores) e 8 GB de memória RAM (1333 MHz, DDR3) com o sistema operacional Windows 8.

As ferramentas utilizadas para a extração de características foram as seguintes:

- *Image Processing Toolbox*;
- LBP: Implementação MATLAB de LBP desenvolvida por Ojala da Universidade de Oulu [33];
- LPQ: Implementação MATLAB de LPQ desenvolvida por Ojansivu da Universidade de Oulu [34];
- AM\_TOOLS: Implementação de *Active Shape Model* e também de *Active Appearance Model* (AAM) desenvolvida por Cootes e Taylor.

As ferramentas utilizadas para a classificação e regressão foram as seguintes:

- LIBSVM versão 3.18: Software integrado para SVM (C-SVC, nu-SVC), SVR (C-SVC, nu-SVC) e estimativa de distribuição (one-class SVM) com interface MATLAB. Suporta problemas de classificação com diversas classes (*multi-class*) [40];
- LIBLINEAR versão 1.94: Software similar ao LIBSVM, porém para classificação linear para os casos onde o conjunto de dados possui um grande número de instâncias e características [41].

Para a etapa de classificação em faixas etárias e estimativa específica da idade utilizando regressão foram utilizadas as seguintes bibliotecas e procedimentos:

- LIBSVM (*C-SVC one-against-one*)
  1. Normalização dos dados para  $([0, 1])$ ;
  2. Utilizado um kernel com função de base radial (*Radial Basis Function - RBF*);
  3. *Grid-search* para encontrar os melhores parâmetros  $C(2^{-1} - 2^3)$  e  $\gamma(2^{-4} - 2)$  utilizando validação cruzada (*cross-validation*) com *5-folds*.
- LIBLINEAR (*one-vs-the rest*)
  1. Normalização dos dados para  $([0, 1])$ ;
  2. Utilizado um kernel linear;
  3. Utilizado o *solver L2-regularized L2-loss support vector classification (dual)* para a classificação;
  4. Utilizado o *solver L2-regularized L2-loss support vector regression (primal)* para a regressão;
  5. *Grid-search* para encontrar o melhor parâmetro  $C(2^{-14} - 2^{14})$  utilizando validação cruzada (*cross-validation*) com *5-folds*.

## 5.1 Extração e análise das características

### 5.1.1 Modelos de aparência ativa

O modelo de aparência ativa é gerado a partir da combinação do modelo estatístico da forma com o modelo estatístico da textura, ou padrões de intensidade. Dois modelos são construídos independentemente, um para o conjunto FG-NET e outro para o MORPH. Para construir o modelo foram utilizadas todas as imagens de cada conjunto anotadas com os 68 *landmarks* que descrevem o formato de diferentes componentes faciais como olhos, sobrancelhas, nariz, boca e contorno do rosto, conforme mostrado na Figura 41. A análise de Procrustes é realizada para alinhar os conjuntos de pontos de referência e construir o modelo estatístico da forma, podendo assim deformar cada exemplo de treinamento para que os pontos correspondam aos da forma média aprendida, obtendo o formato livre da imagem. As intensidades do formato livre da imagem são normalizadas a partir de uma transformação linear para que a técnica PCA possa ser aplicada para construir o

modelo estatístico da textura. Finalmente, uma correlação entre o modelo do formato e o da textura é aprendida para gerar um modelo combinado da aparência, conforme descritos na seção 3.2.1. Para lidar com as variações de resolução presentes nos conjuntos de dados é construído um modelo multiresolução baseado em uma pirâmide Gaussiana [42] que permite a análise multiresolução levando a um aumento na velocidade e robustez da técnica. Para cada nível da pirâmide é construído um modelo de textura independente, sendo amostrado um número de pixels proporcional a área da região alvo da imagem, no caso deste trabalho, foram amostrados aproximadamente 20.000 valores de intensidade da imagem facial a partir do modelo da forma média aprendido. Portanto, o primeiro nível ( $n_0$ ) possui 20.000 pixels, o segundo nível ( $n_1$ ) tem a metade do  $n_0$ , ou seja, 10.000 pixels e o nível 2 ( $n_2$ ) possui a metade do  $n_1$ , sendo 5.000 pixels amostrados. Para os níveis  $n_1$  e  $n_2$  é aplicado uma suavização com um filtro gaussiano antes da subamostragem.

Para extrair as características relacionadas a textura da face o AAM foi configurado para utilizar o algoritmo de envoltória convexa (*convex hull*) [43] nos componentes faciais utilizando triangulação (Figura 41). A envoltória convexa de um conjunto finito de pontos é o menor polígono (2D) / poliedro (3D) convexo que contém todos os pontos dados, ou seja, é a interseção de todos os polígonos/poliedros que contém os pontos. O nível de truncamento para os detalhes que representam o formato, textura e o modelo combinado foi configurado para 95%. Isso significa que o modelo resultante descreve 95% das variações aparentes nos conjuntos de treinamentos.

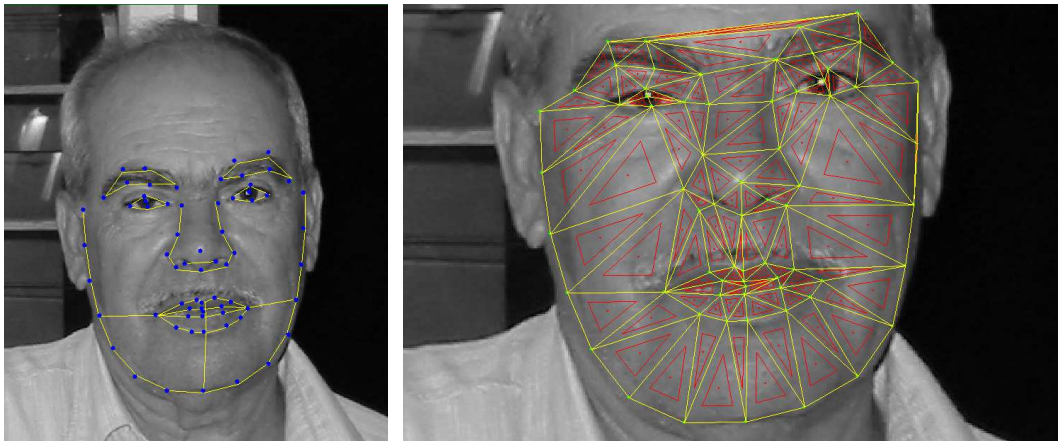


Figura 41: Os 68 *landmarks* descrevendo os formatos de componentes faciais e a triangulação para a descrição da textura, respectivamente.

Devido o AAM ter sido desenvolvido para modelamento facial características como variações de identidade são melhores representadas do que características relacionadas ao envelhecimento. Portanto, algumas características relevantes como rugas e

texturas da pele são removidas após a redução da dimensionalidade das características pelo PCA. Isso pode ser percebido na Figura 42 que mostra uma comparação da imagem original, a imagem reconstruída pelo parâmetros combinados do AAM e o modelo residual sendo a diferença absoluta entre a imagem original e a reconstruída. A imagem reconstruída não representa suficientemente as características relacionadas ao envelhecimento. Por exemplo, as rugas nos cantos dos olhos e as linhas nasolabiais do indivíduo da Figura 42 não são bem representadas pelo modelo. Baseando-se na Figura 42 é possível concluir que técnicas para a extração de características locais como rugas e outras marcas na pele também são necessárias para a estimação da idade uma vez que o AAM não as representa detalhadamente. O vetor de características resultante do modelo combinado após a aplicação do PCA possui apenas 62 características mostrando o poder da redução de dimensionalidade aplicada pelo PCA.

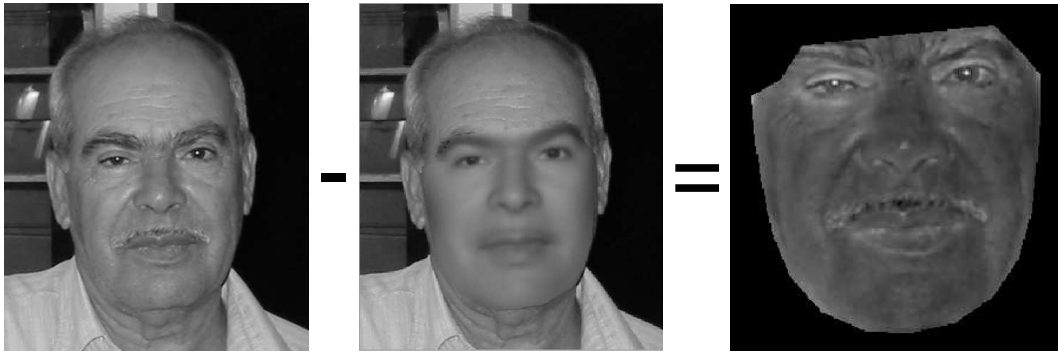


Figura 42: Imagem original, imagem reconstruída a partir dos parâmetros combinados do AAM e o modelo residual gerado pela diferença absoluta entre a imagem original e reconstruída.

A Figura 43 mostra os resultados a partir da sintetização do modelo facial utilizando as 5 primeiras características combinadas em ambas as direções do desvio padrão aprendido, assim é possível ter uma ideia de quais são as mudanças que cada característica causa no modelo facial. A primeira característica (linha 1) mostra a face média aprendida pelo AAM. É possível observar que as 4 primeiras características cobrem principalmente a posição da cabeça e variações de luminosidade devido a grande variação de imagens não controladas na base de dados. Já a característica 5 contém mais informações relacionada a idade, como o formato facial e textura, justificando o uso do AAM como características globais para a estimativa da idade.

A Figura 44 mostra alguns exemplos de algumas faces reconstruídas a partir do modelo AAM combinado da base de dados FG-NET. Pode-se perceber que as faces reconstruídas são muito parecidas com as faces originais mostrando a eficiência do AAM



Figura 43: Variações das 5 primeiras características combinadas do AAM em ambas as direções começando com a face média (1), sempre mostrada ao centro (FG-NET).

para o modelamento de faces.

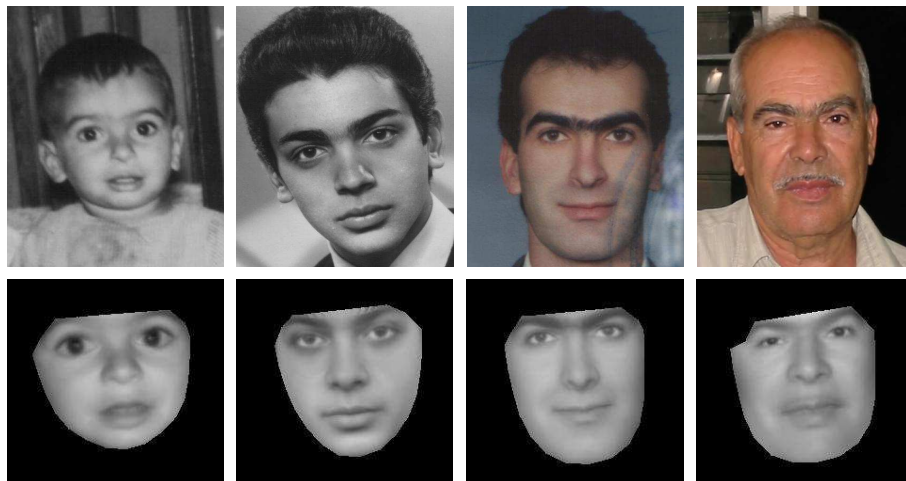


Figura 44: Exemplos de faces reconstruídas a partir do modelo AAM combinado (2, 17, 28 e 69 anos respectivamente) (FG-NET).

### 5.1.2 Gabor *wavelets*

As rugas faciais fornecem informações importantes relacionadas a estimaco etria, por isso, foram consideradas nessa pesquisa. Criadas pelo movimento repetitivo dos msculos da face e expresses formam-se perpendicularmente  direco dos msculos faciais. Por exemplo, as rugas da testa tem uma direco horizontal enquanto as rugas nasolabiais so diagonais. Conseqentemente, rugas faciais possuem caractersticas nicas que dependem da direco dos msculos, a frequncia de uso e os efeitos da gravidade.

Os filtros de Gabor so utilizados nessa pesquisa para a deteco de linhas, e como uma imagem facial pode conter linhas, ou rugas, com diferentes espessuras  necessrio construir filtros com diferentes fatores de escala ( $a$ ). Tambm, devido uma imagem possuir rugas em diferentes orientaes, os filtros so gerados em diferentes ângulos ( $\theta$ ). Para extrair essas caractersticas foi utilizado um conjunto de filtros de Gabor com diferentes parmetros de orientao e escala para serem aplicados nas 11 regies de interesse de forma a capturar as rugas faciais em diversas escalas e orientaes. A Figura 45 mostra os filtros e magnitudes de Gabor para 8 orientaes (colunas) e 5 escalas (linhas) utilizados para extrair caractersticas relacionadas a rugas faciais. Utilizando todos esses 40 filtros  possvel extrair uma grande quantidade de informaes relacionadas as intensidades e orientaes das rugas faciais utilizando a mdia e varincia da resposta em magnitude de cada regio de interesse, pois representam tanto a intensidade quanto a quantidade das rugas. Tambm, foi analisado um conjunto de filtros contendo 6 orientaes e 4 escalas diferentes, totalizando 24 filtros de Gabor, onde obteve um desempenho similar em termos de *MAE* na estimaco etria quando comparado com a utilizao de 40 filtros, conforme resultados mostrados na Tabela 23 da seo 5.4, porm, devido possuir uma quantidade menor de filtros, o custo computacional  reduzido, sendo uma vantagem a utilizao da configurao com 6 orientaes e 4 escalas.

Para a anlise da tcnica Gabor *wavelets* foi utilizada a regio da testa de um indivduo de 2 e 69 anos, pois  possvel perceber facilmente as linhas horizontais causadas pelas rugas. Tambm, a regio periorbital (canto do olho) onde possui rugas em vrias direces de acordo com a idade (ps de galinha). A Figura 46 mostra a regio da testa de um indivduo de 2 e 69 anos utilizada para a aplicao dos filtros de Gabor descritos na Figura 45. As magnitudes dos filtros de Gabor, que so utilizadas como caractersticas faciais, e seus valores mdios da regio da testa de um indivduo de 2 e 69 anos so mostrados na Figura 47. Devido a testa de um beb possuir uma pele suave percebe-se que a resposta em magnitude dos filtros possuem pouca intensidade por ter poucas linhas

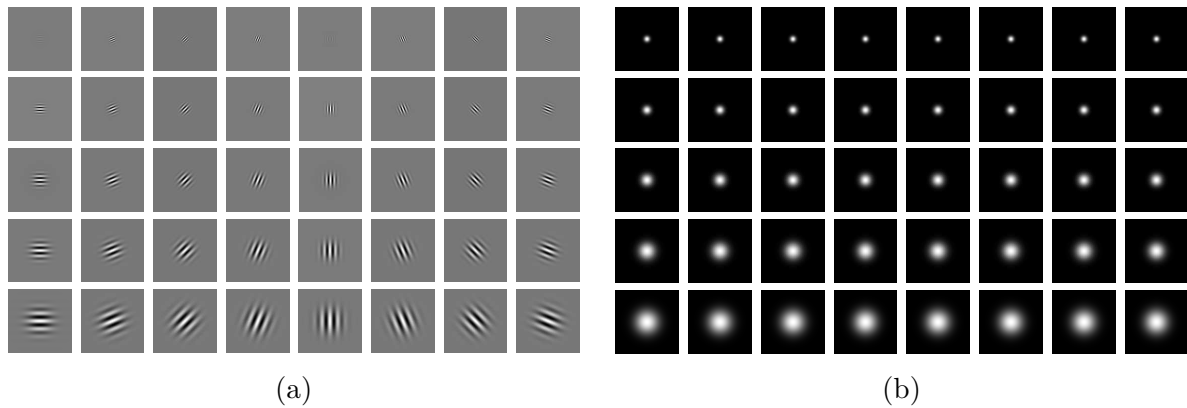


Figura 45: Filtros e magnitudes de Gabor para 8 orientações (colunas) e 5 escalas (linhas): (a) filtros e (b) magnitudes.

faciais. Enquanto a região da testa do indivíduo de 69 anos nessa região possui uma pele rugosa com linhas finas e espessas em diversas direções percebendo-se que a resposta em magnitude dos filtros possuem bastante intensidade, ou conteúdo frequencial, em várias orientações e escalas. Em geral, os valores médios de magnitude da região da testa do indivíduo de 69 anos são maiores, como é possível verificar na Figura 47.

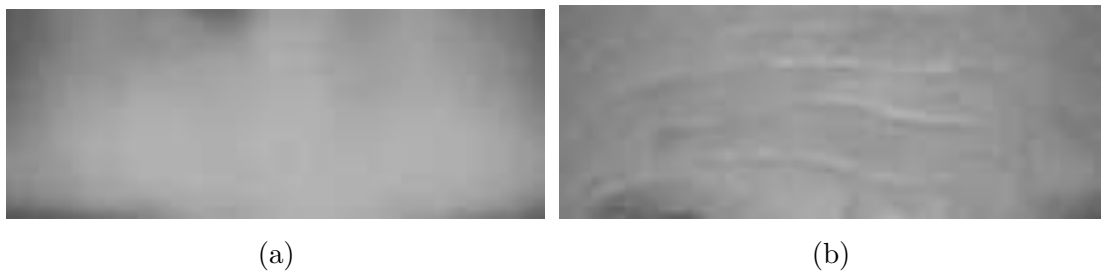


Figura 46: Regiões da testa utilizadas para a aplicação dos filtros de Gabor (FG-NET): (a) 2 anos e (b) 69 anos de idade.

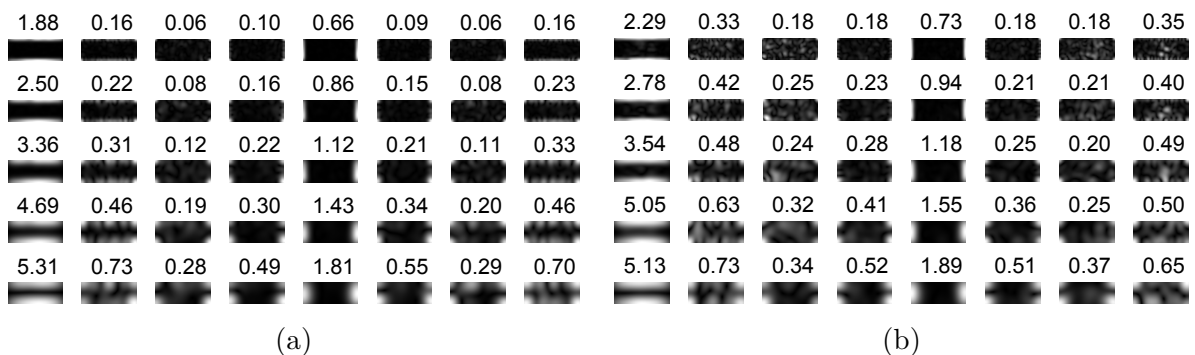


Figura 47: Magnitudes dos filtros de Gabor e seus valores médios para 8 orientações (colunas) e 5 escalas (linhas) das regiões da testa: (a) 2 anos e (b) 69 anos de idade.

A Figura 48 mostra a região periorbital de um indivíduo de 2 e 69 anos utilizada para a aplicação dos filtros de Gabor descritos na Figura 45. As magnitudes dos filtros

de Gabor e seus valores médios da região periorbital de um indivíduo de 2 e 69 anos são mostrados na Figura 49. A pele de um bebê nessa região normalmente possui uma textura suave. Idealmente, a resposta em magnitude dos filtros deveriam possuir pouca intensidade por ter poucas linhas faciais. Porém, existem linhas ruidosas devido ao fato da região de interesse ter sido cortada pegando uma região da orelha e cabelo. Já a pele do indivíduo de 69 anos nessa região possui linhas faciais em diversas direções, com isso, é possível perceber que a resposta em magnitude dos filtros possuem bastante intensidade em várias orientações e escalas. Na região delimitada em vermelho da Figura 49 é possível perceber que os valores de intensidade das magnitudes são maiores que os da pele do bebê, pois foram utilizados filtros com as mesmas orientações e escalas das rugas do canto do olho do indivíduo de 69 anos, onde as linhas faciais saem do canto do olho e vão para baixo.

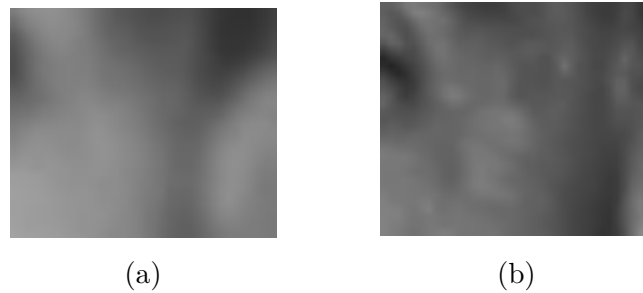


Figura 48: Regiões periorbital utilizada para a aplicação dos filtros de Gabor (FG-NET): (a) 2 anos e (b) 69 anos de idade.

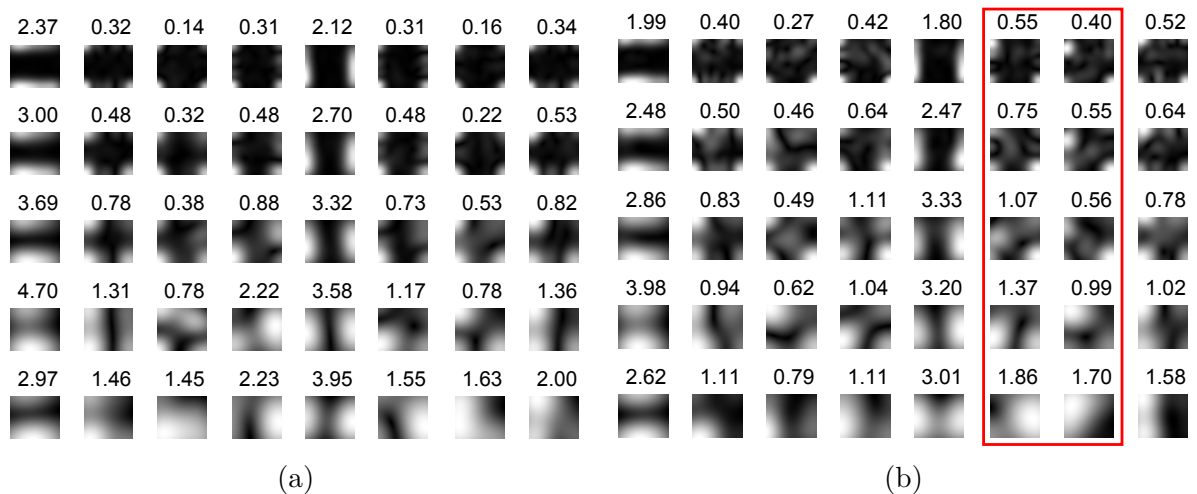


Figura 49: Componentes reais dos filtros de Gabor para 8 orientações (colunas) e 5 escalas (linhas) das regiões periorbital: (a) 2 anos e (b) 69 anos de idade.

Neste trabalho, a técnica Gabor *wavelets* foi aplicada nas regiões de interesse com os mesmos pré-processamentos descritos na seção 4.2, não sendo considerada uma



equalização no contraste da imagem que poderia destacar melhor a rugas faciais. Porém, mesmo sem o balanceamento na iluminação, os resultados foram satisfatórios, conforme mostra a Tabela 23.

### 5.1.3 *Local binary patterns*

As características locais da textura das 11 regiões de interesse foram extraídas a partir do operador LBP utilizando variações em seus parâmetros, como a configuração da vizinhança circular que utiliza  $n$  pixels vizinhos (e.g., 8 pixels) distantes de  $m$  pixels (e.g., 2 pixels de raio) do pixel central. Foram utilizados, por exemplo, as configurações  $LBP_{n,m}^{u2}$ , onde  $n$  assume 8 ou 16 pixels e  $m$  pode variar de 1 a 3 pixels, sendo possível fazer diversas combinações com o operador. A nomenclatura  $u2$  significa que foram utilizados padrões uniformes, o qual foi utilizado em todo o trabalho, para reduzir os efeitos causados por ruídos nas estruturas locais de textura. Contabilizando os padrões uniformes e não-uniformes têm-se 256 padrões, ou códigos, porém, quando utilizado apenas padrões uniformes, têm-se um total de 58 padrões mais 1 intervalo para agregar os padrões não-uniformes, totalizando 59 códigos. Pelo fato de existir uma simetria facial foram utilizados apenas as regiões de interesse do lado direito do rosto para as análises. A Figura 50 mostra alguns padrões uniformes detectados pelo operador LBP como término de uma linha, bordas, pontos escuros e claros, pontiagudo etc. O valor 1 é representado como círculo branco e 0 como preto.

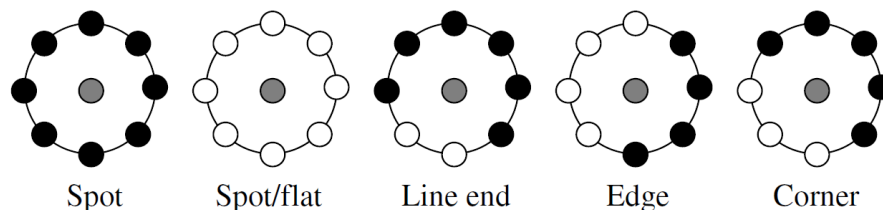


Figura 50: Diferentes padrões detectados pelo operador LBP [13].

A Figura 51 mostra o histograma dos códigos  $LBP_{8,2}^{u2}$  para a região das linhas horizontais da testa. Foi utilizado um exemplo para cada faixa etária: bebê de 2 anos, jovem de 17 anos, adulto de 28 anos e idoso de 69 anos. A pele da região da testa de um bebê de 2 anos é normalmente uma pele lisa e suave como mostra o código 58 que se mostra muito proeminente, o qual representa um ponto escuro com os pontos circundantes mais brilhantes, ou seja, uma região com textura suave, conforme indicado no histograma. Consequentemente, com poucas manchas (*spots*), onde todos os pontos circundantes possuem um valor mais baixo que o pixel central, como indicadas no histograma pelo código 1

(*manchas*). Alguns códigos relacionados a bordas são devido à região de interesse ter sido cortada pegando um pouco da região das sobrancelhas e cabelo, indicados pela delimitação em vermelho na figura. O histograma do indivíduo de 17 anos possui mais códigos relacionados a manchas e também uma predominância do código 59 que representam padrões não-uniformes os quais definem várias curvaturas de linhas faciais. No indivíduo de 28 anos o histograma é similar ao do indivíduo de 17 anos. Porém, alterando alguns códigos que são derivados de ruídos causados pelas sobrancelhas em ambos os histogramas. No histograma do idoso de 69 anos é possível perceber códigos relacionados a manchas causadas pelo envelhecimento. Vários códigos relacionados a bordas (e.g., 11 e 15) e cantos (e.g., 7, 10, 15 e 25) também são observados, causadas pelas linhas horizontais na testa. E também, um número maior do código 59 que representam padrões não-uniformes os quais definem várias curvaturas de linhas faciais.

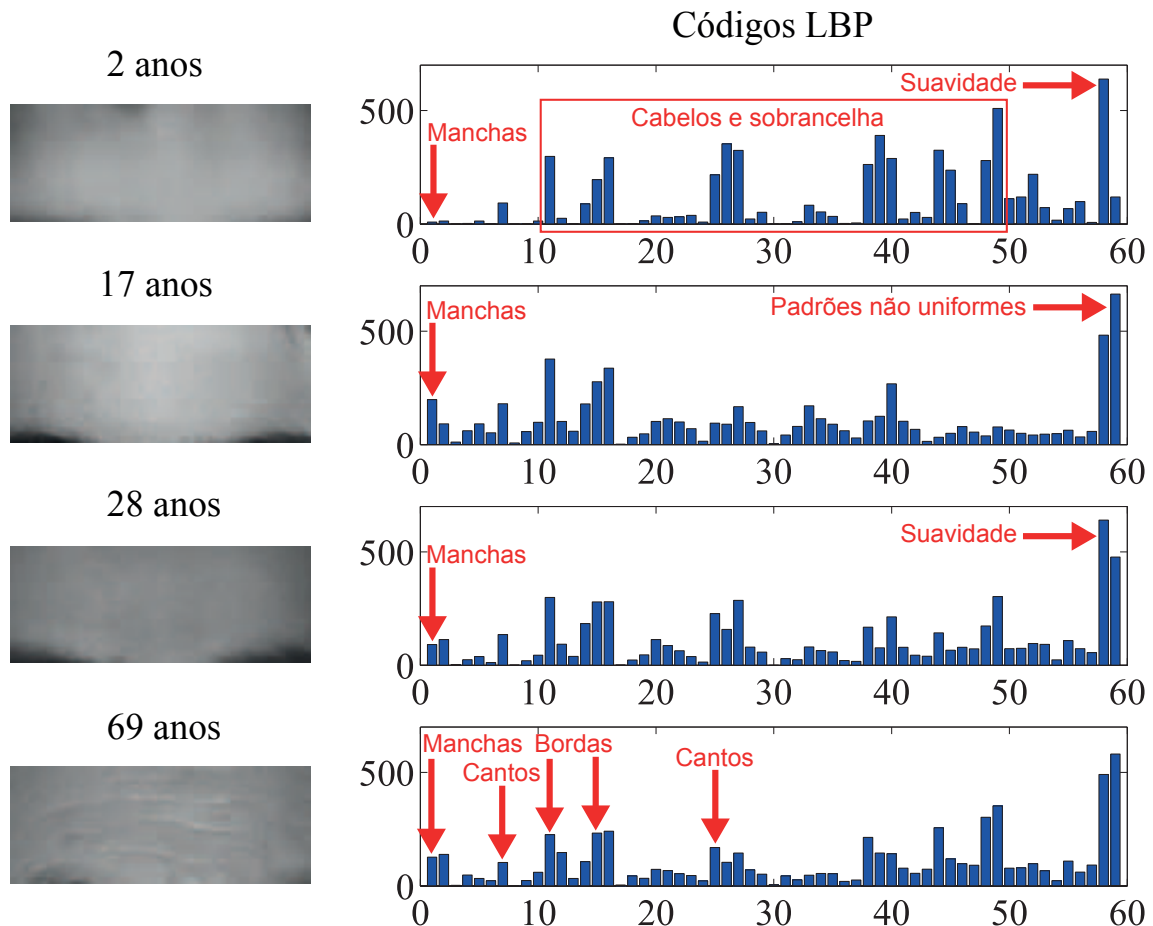


Figura 51: Histogramas dos códigos  $LBP_{8,2}^{u_2}$  para a região das linhas horizontais da testa para 2, 17, 28 e 69 anos respectivamente (FG-NET).

A Figura 52 mostra o histograma dos códigos  $LBP_{8,2}^{u_2}$  para a região do canto do olho dos mesmos indivíduos descritos anteriormente. A região do canto dos olhos é uma

das regiões com maior número de rugas sendo uma região importante para a extração de características relacionadas ao envelhecimento, como por exemplo, as rugas pés de galinha. A pele dos cantos dos olhos de um bebê de 2 anos provavelmente não terá pés de galinha, será uma pele mais lisa com poucas manchas claras, como é possível perceber no histograma através do código 1 (indicado na figura). Os códigos 20, 26 e 27, indicados no histograma do bebê de 2 anos, representam pontiagudos e bordas na imagem que são ruídos causados pelo limite da pele com o cabelo e região da orelha. Os histogramas dos indivíduos de 17 anos e 28 anos possuem poucos códigos relacionados a manchas como mostra o código 1 indicado no histograma. Assim como o bebê, a pele de uma pessoa de 17 anos e 28 anos normalmente é suave sem rugas periorbital em expressões neutras, porém existem alguns ruídos causados pelas bordas do canto do olho. No histograma do indivíduo de 69 anos os pés de galinha já são destacados conforme descrevem os códigos 7, 19 e 33 que representam bordas e terminos de linhas das rugas faciais. Também, um número maior de códigos 1 podem ser observados os quais representam manchas na pele. No Apêndice A são mostradas as análises do canto direito da boca, linhas da bochecha direita, linhas do queixo, topo do nariz e linhas nasolabial dos mesmos indivíduos utilizados nas análises anteriores.

#### 5.1.4 *Local phase quantization*

Nos experimentos utilizando a técnica LPQ os descritores foram extraídos utilizando-se a janela  $M$  (*winSize*) de  $5 \times 5$  com os parâmetros  $a = 1/3$  (alpha na abordagem STFT) e  $\rho = 0.9$  (coeficiente de correlação). A janela de  $5 \times 5$  foi escolhida pelo fato de ter obtido o melhor resultado em termos de *MAE* quando aplicado isoladamente como uma característica LPQ local, conforme mostra a Tabela 23 no final da seção 5.4.

Os 256 códigos LPQ resultantes são derivados da computação de cada pixel da região de interesse utilizando a janela  $M$ . O histograma resultante é então utilizado para a classificação das características faciais. É possível perceber em todos os histogramas mostrados nas Figuras 53 e 54 que os valores dos códigos LPQ são geralmente maiores para indivíduos jovens (2, 17 e 28 anos), pois possuem graus de orientação de fases similares, normalmente em um mesmo sentido devido a possuírem uma textura de pele mais suave. No caso do indivíduo de 69 anos a textura da pele possui rugas em diversas orientações, como é possível perceber pelo gráfico de orientações das figuras, onde os vetores unitários que demonstram a orientação das características estão em diversos sentidos. Também, conforme a pele vai ficando com uma textura mais envelhecida o histograma de códigos

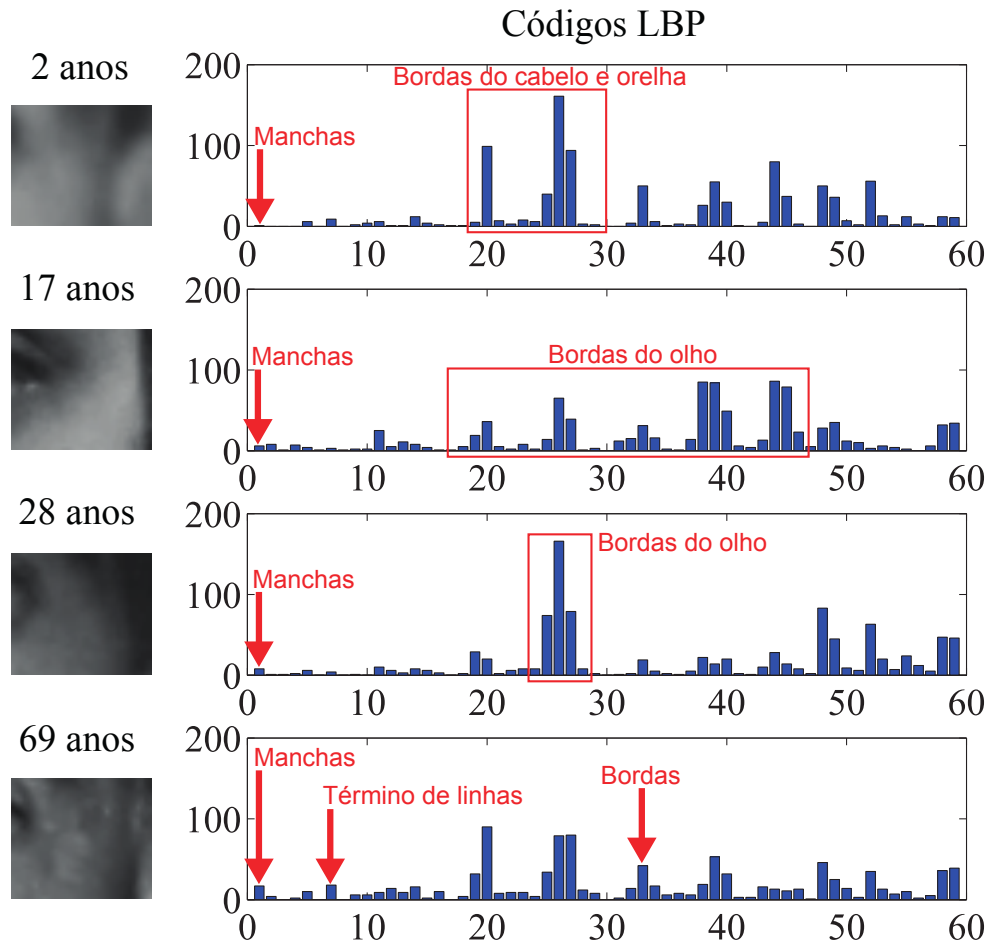


Figura 52: Histogramas dos códigos  $LBP_{8,2}^{u_2}$  para a região das linhas periorbital direitas para 2, 17, 28 e 69 anos respectivamente (FG-NET).

LPQ vai tendo uma distribuição mais uniforme com valores baixos devido a possuir mais graus de orientação das fases causado pelas rugas. Alguns códigos de fases não desejados, principalmente vistos e indicados nos histogramas dos indivíduos de 17 e 28 anos, são causadas devido a região da testa pegar parte das sobrancelhas e a região das linhas periorbital pegar uma parte do canto do olho, porém, tais partes indesejadas nas regiões de interesse foram desconsideradas neste trabalho. No Apêndice B são mostradas as análises do canto direito da boca, linhas da bochecha direita, linhas do queixo, topo do nariz e linhas nasolabial dos mesmos indivíduos.

As técnicas de extração de características analisadas nas seções anteriores mostraram ser muito discriminantes para o envelhecimento facial. Os modelos de aparência ativa contêm informações relevantes relacionadas ao formato facial, porém descrições importantes sobre a textura são perdidas após a aplicação do PCA. No entanto, são necessárias outras técnicas para a extração de características locais. A utilização de Gabor *wavelets*

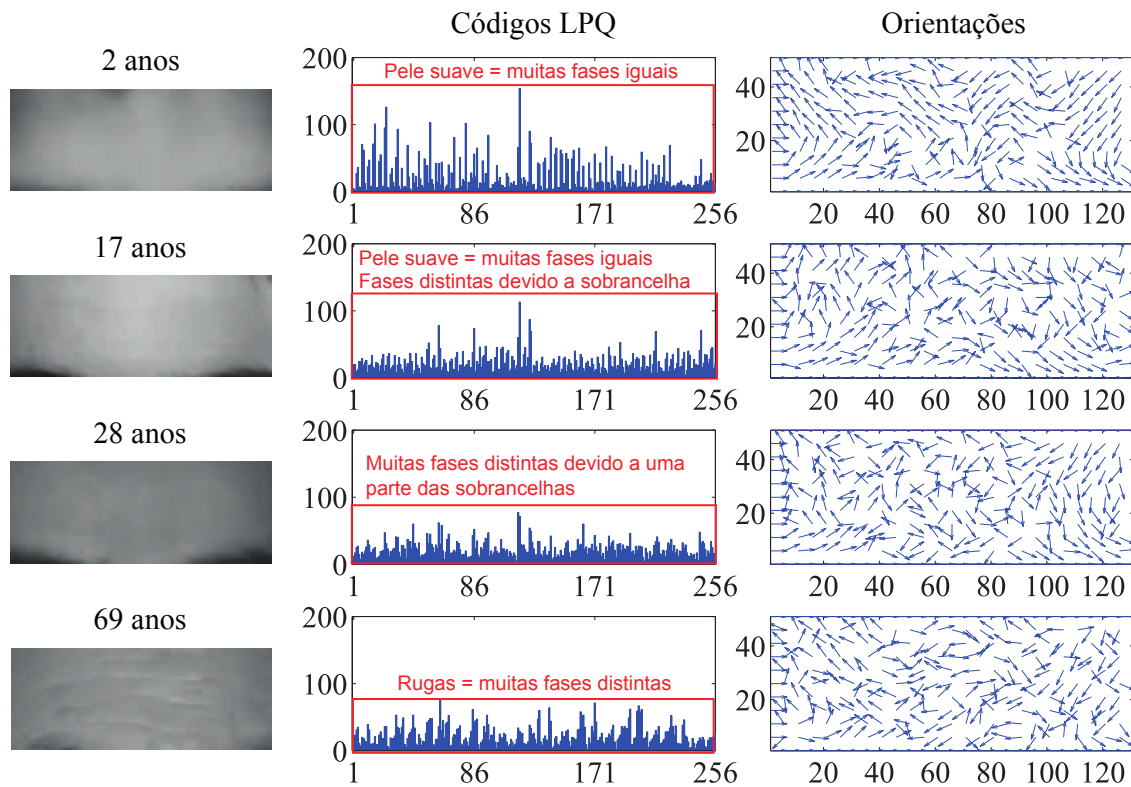


Figura 53: Histogramas dos códigos  $LPQ_{5 \times 5}$  para a região das linhas horizontais da testa e suas respectivas orientações para 2, 17, 28 e 69 anos respectivamente (FG-NET).

para a detecção de rugas faciais tem se mostrado eficiente quando as rugas estão bem visíveis na imagem. O descritor de textura LBP também foi bastante discriminativo capturando principalmente informações como bordas e manchas na pele. Já o operador LPQ se mostrou muito poderoso na descrição da textura com a utilização de informações de fase locais, possuindo histogramas distintos entre a pele de um bebê e idoso. A etapa de extração das características é a mais importante para o desenvolvimento do método de estimação etária, pois quanto mais discriminantes forem as características maior será o desempenho da classificação hierárquica da idade.

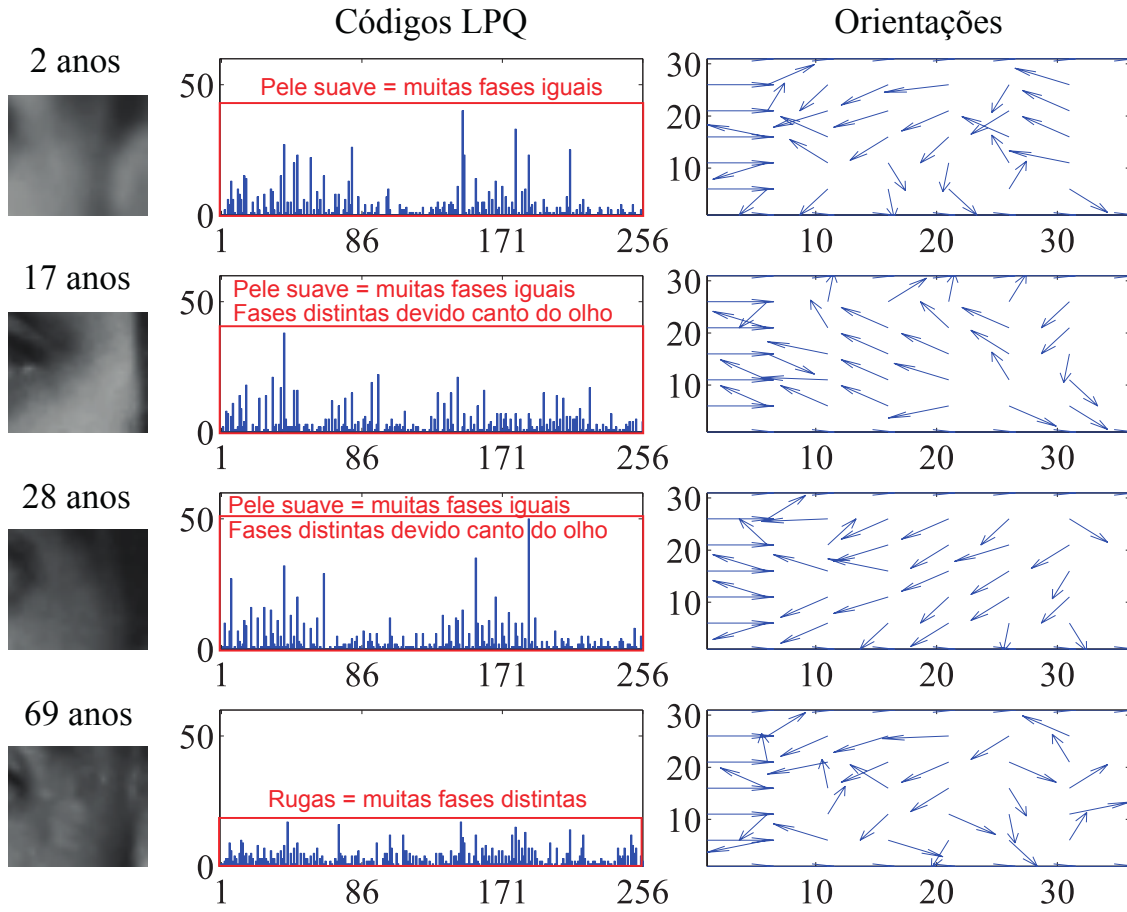


Figura 54: Histogramas dos códigos  $LPQ_{5 \times 5}$  para a região das linhas periorbital direita e suas respectivas orientações para 2, 17, 28 e 69 anos respectivamente (FG-NET).

## 5.2 Métricas de avaliação

O desempenho do método proposto foi avaliado através do erro médio absoluto ( $MAE$ ) e do escore acumulativo ( $CS$ ) [2]. O  $MAE$  é a média dos erros absolutos entre a idade estimada ( $a$ ) e a idade real (*ground truth*) ( $\hat{a}$ ) definida como,

$$MAE = \frac{\sum_{i=1}^N |\hat{a} - a|}{N}, \quad (5.1)$$

onde  $N$  é o total de imagens utilizadas para teste. O  $CS$  permite uma comparação de desempenho em diferentes níveis de erros absolutos e é expresso como,

$$CS(l) = \frac{N_{e \leq l}}{N} \times 100\%, \quad (5.2)$$

onde  $N_{e \leq l}$  é o total de imagens utilizadas para teste onde a idade estimada possui um

erro absoluto menor ou igual a  $l$  anos.

### 5.3 Protocolos de avaliação

Para todos os experimentos realizados neste trabalho foram utilizado 75% dos dados para treinamento e 25% para teste para ambos os conjunto de dados, FG-NET *Aging* e MORPH *Album 2*, assim como para a combinação dos conjuntos. Também, os classificadores e regressores foram configurados para fazer uma validação cruzada utilizando 5 *folds* e com um busca pelos melhores parâmetros SVM e SVR através do método *grid-search*.

Devido o conjunto de dados FG-NET *Aging* possuir poucos dados (1.002 imagens), o protocolo *leave-one-person-out* (LOPO) também foi utilizado para avaliar o desempenho do método proposto. O protocolo LOPO considera todas as amostras de um único indivíduo para teste enquanto o restante é usado para o treinamento. Esse processo faz com que cada indivíduo seja testado uma vez isoladamente, resultando em 82 *folds*, pois o conjunto FG-NET possui 82 indivíduos. Este esquema garante que uma pessoa não esteja no conjunto de teste e treinamento ao mesmo tempo, portanto, o classificador não pode aprender características individuais, eliminando assim a dependência dos dados nos resultados do experimento.

### 5.4 Conjunto de dados FG-NET *Aging*

Nesta seção são apresentados os resultados da estimativa da idade no conjunto de dados FG-NET utilizando métodos de estimação *single-level*, hierárquica com margens fixas e com regiões sobrepostas. São utilizados diversos conjuntos de características locais, globais, assim como a combinações de ambos afim de encontrar a melhor representação do envelhecimento facial e também o método de estimação etária com o melhor desempenho.

#### 5.4.1 Estimativa *Single-level*

O método de estimativa *single-level* utiliza apenas uma etapa de regressão SVR para estimar a idade, e foi avaliado para poder comparar com métodos hierárquicos que utilizam uma quantidade maior de regressores selecionados de acordo com os resultados de uma etapa classificatória. O regressor SVR unitário foi treinado em toda a faixa etária da base de dados FG-NET, ou seja, na faixa de 0 a 69 anos. A distribuição dos dados de treinamento e teste utilizada é mostrada na Tabela 5.

Tabela 5: Distribuição dos dados de treinamento e teste para a estimativa *single-level* (FG-NET).

SVR	Faixa etária	Treinamento	Teste
1	0-69	752	250

Os cinco melhores resultados em termos de  $MAE$  para o método *Single-level* são mostrados na Tabela 6. A combinação das características globais e locais AAM+GW<sub>8,5</sub> e AAM+LPQ<sub>9×9</sub> foram as que tiveram melhor desempenho, sendo um  $MAE$  de 5,31 anos em ambas as combinações. No Apêndice D é mostrada uma tabela contendo todos os resultados experimentais do método *single-level* em diversos conjuntos de características. O desempenho geral do estimador *single-level* foi de 6,70 anos, mostrando-se satisfatório especialmente pela simplicidade da implementação, pois utiliza apenas um regressor. Porém, se as características faciais extraídas são treinadas em uma única faixa etária, características distintas relacionadas a uma determinada faixa não são levadas em conta, não conseguindo fazer uma estimativa precisa da idade. Portanto, a estimativa hierárquica da idade com regiões sobrepostas, que serão descritas posteriormente, possuem um melhor desempenho.

Tabela 6: Cinco melhores resultados para o método *Single-level* (FG-NET).

Características	<i>Single-level</i> $MAE$
AAM	5,64
AAM+GW <sub>6,4</sub>	5,49
AAM+GW <sub>8,5</sub>	<b>5,31</b>
AAM+LPQ <sub>7×7</sub>	5,51
AAM+LPQ <sub>9×9</sub>	<b>5,31</b>

## 5.4.2 Estimativa hierárquica com margens fixas

A estimativa hierárquica com margens fixas é dividida em duas etapas: Classificação em faixas etárias e estimativa específica da idade. Por não possuir regiões sobrepostas entre as faixas etárias é chamado de margens fixas.

### 5.4.2.1 Classificação em faixas etárias

Para a etapa de classificação em faixas etárias foram definidas 4 classes. As faixas de cada classe foram definidas de acordo com a distribuição dos dados da FG-NET



para que cada classe tenha dados suficientes para treinamento e teste de uma forma bem distribuída. Também, as classes foram definidas de acordo com os grupos etários: crianças, jovens, adultos e idosos. As faixas etárias e a distribuição dos dados de treinamento e teste para cada classe são mostradas na Tabela 7.

Tabela 7: Faixas etárias e distribuição dos dados de treinamento e teste para cada classe (FG-NET).

Classe	Faixa etária	Treinamento	Teste
1	0-13	385	128
2	14-21	175	58
3	22-39	140	47
4	40-69	52	17
Total	0-69	752	250

A Figura 55 mostra o desempenho da classificação em cada conjunto de características utilizando o kernel RBF através da ferramenta LIBSVM e também o kernel linear através da ferramenta LIBLINEAR. A maior taxa de acerto atingida utilizando kernel RBF foi de 70% com a combinação de características globais e locais AAM+GW<sub>6,4</sub>. Utilizando o kernel linear a maior taxa de acerto foi atingida com a combinação AAM+LPQ<sub>(9×9)</sub> tendo uma taxa de 68,8% de instâncias classificadas corretamente. Em geral, o desempenho da classificação utilizando SVM com kernel linear é mais alto, como mostra a curva respectiva no gráfico.

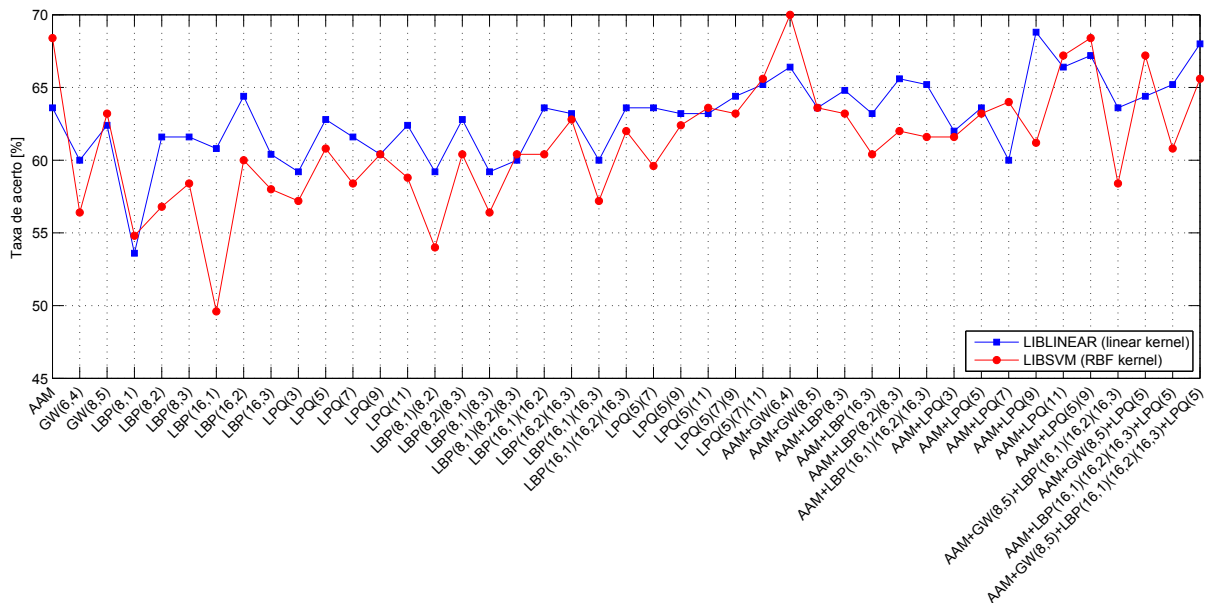


Figura 55: Comparação do desempenho da classificação utilizando kernel RBF (LIBSVM) e linear (LIBLINEAR) em diferentes conjuntos de características (FG-NET).

A matriz de confusão da classificação em faixas etárias utilizando a combinação de características globais e locais AAM+LPQ<sub>(7×7)</sub>, que obtiveram o melhor desempenho da estimativa hierárquica da idade com margens fixas (Tabela 8) é mostrada na Figura 56. A taxa de acerto geral da classificação foi de 60%. A classe 1 (0-13 anos) foi a que obteve a melhor taxa de acerto na classificação, 92,2%, tendo um erro de apenas 7,8% onde parte das instâncias de teste foram classificadas dentro das classes 2 (14-21 anos) e 3 (14-21 anos). A classe 2 obteve um erro de 70,7% devido a classificar erroneamente a maioria das instâncias de teste na classe 1. A classe 3 teve um erro de 70,2%, classificando erroneamente algumas instâncias de teste nas classes 1 e 2. Isso se deve ao fato de existirem indivíduos da classe 3 (22-39 anos) que podem ser classificados dentro das classes 1 ou 2 por possuírem características semelhantes como poucas rugas e sem marcas na pele devido a fatores genéticos, estilos de vida, saúde etc. Devido possuir poucos dados de treinamento e teste a classe 4 (40-69 anos) teve o pior desempenho, 94,1% das instâncias foram classificadas erroneamente em outras classes. A matriz de confusão mostra claramente a tendência da base FG-NET em ser mais precisa na faixa etária de 0 a 13 anos. As confusões também ocorreram devido as margens das faixas etárias serem fixas (*hard boundaries*), onde um indivíduo de 14 anos (classe 2), por exemplo, pode ser facilmente classificado dentro da classe 1 que possui uma faixa etária de 0 a 13 anos.

		Taxa de acerto				
Classe predita	1	118 47.2%	9 3.6%	1 0.4%	0 0.0%	92.2% 7.8%
	2	36 14.4%	17 6.8%	5 2.0%	0 0.0%	29.3% 70.7%
	3	15 6.0%	18 7.2%	14 5.6%	0 0.0%	29.8% 70.2%
	4	4 1.6%	1 0.4%	11 4.4%	1 0.4%	5.9% 94.1%
		1	2	3	4	
		Classe verdadeira				

Figura 56: Matriz de confusão da classificação em faixas etárias utilizando a combinação de características globais e locais AAM+LPQ<sub>(7×7)</sub> (FG-NET).

Tabela 8: Resultados de diferentes métodos de estimação etária em diferentes conjuntos de características, mostrando também a dimensão original das características e depois da redução da dimensionalidade utilizando PCA (FG-NET).

Características	Dimensão Orig./PCA	Classificação Taxa de acerto [%]	Regressão <i>MAE</i>	<i>Single-level MAE</i>	Hierárquico <i>MAE</i>	Hierárquico regiões sobrepostas fixas <i>MAE</i>	Hierárquico regiões sobrepostas flexíveis <i>MAE</i>
AAM	-/62	63,6	2,94	5,64	5,91	5,11	5,65
GW <sub>8,5</sub>	8331/180	62,4	3,27	6,44	7,54	7,40	7,44
LBP <sub>16,3</sub>	2673/241	60,4	3,66	7,11	6,42	6,23	6,12
LPQ <sub>7×7</sub>	2816/197	61,6	3,64	6,62	7,31	7,35	6,79
LBP <sub>(16,1)(16,2)(16,3)</sub>	564/206	63,6	3,47	7,38	6,16	6,28	6,03
LPQ <sub>(5×5)(9×9)</sub>	395/212	63,2	3,34	6,42	5,44	5,49	4,86
AAM+GW <sub>6,4</sub>	299/54	66,4	<b>2,61</b>	5,49	6,02	5,98	6,06
AAM+GW <sub>8,5</sub>	242/58	63,6	2,86	<b>5,31</b>	5,61	5,54	5,25
AAM+LBP <sub>16,3</sub>	303/191	63,2	3,54	6,25	5,92	6,25	5,32
AAM+LBP <sub>(16,1)(16,2)(16,3)</sub>	268/265	65,2	3,20	6,93	5,42	6,53	5,17
AAM+LPQ <sub>7×7</sub>	259/158	60	3,51	5,51	5,31	5,30	<b>4,50</b>
AAM+GW <sub>8,5</sub> +LBP <sub>(16,1)(16,2)(16,3)</sub>	448/297	63,6	3,25	6,51	5,28	6,06	4,86
AAM+GW <sub>8,5</sub> +LPQ <sub>5×5</sub>	442/137	64,4	3,22	5,78	5,36	<b>4,85</b>	5,12
AAM+LBP <sub>(16,1)(16,2)(16,3)</sub> +LPQ <sub>5×5</sub>	468/329	65,2	3,30	6,38	5,33	4,96	4,95
AAM+GW <sub>8,5</sub> +LBP <sub>(16,1)(16,2)(16,3)</sub> +LPQ <sub>5×5</sub>	648/151	68	3,21	5,88	5,17	5,79	5,32

#### 5.4.2.2 Estimativa específica da idade

Para a etapa da estimativa específica da idade foram utilizadas as mesmas faixas etárias e distribuições de dados descritas anteriormente na etapa da classificação (Tabela 7) para definir os regressores SVR. De acordo com a classificação da faixa etária de um indivíduo, um regressor específico é selecionado para estimar a idade. A partir dessa abordagem hierárquica, a combinação de características que obteve o melhor desempenho para a estimativa com margens fixas foi AAM+LPQ<sub>(5×5)</sub> com um *MAE* de 5,13 anos. Considerando o desempenho individual de cada regressor, o SVR 1, treinado na faixa etária de 0-13 anos obteve um *MAE* de 4,93 anos, o SVR 2 (14-21 anos) 5,79 anos, o SVR 3 (22-39 anos) 9,03 anos e o SVR 4 obteve um *MAE* de 0,78 anos. Na Figura 57 são mostrados os resultados da estimativa hierárquica (linha *Hierárquico*) em diferentes conjuntos de características. Na Tabela 8 são mostrados os resultados numéricos de alguns conjuntos de características utilizados para a estimativa hierárquica da idade (coluna *Hierárquico MAE*). Também é mostrado na coluna *Regressão MAE* os resultados médios dos regressores considerando que a classificação fosse 100% correta. Nesse caso, a melhor

combinação foi a AAM+GW<sub>6,4</sub> com um *MAE* de 2,61 anos. Com esses resultados é possível concluir que grande parte dos erros são causados na etapa de classificação, no caso de uma classificação perfeita o erro seria muito baixo, como é possível perceber na coluna *Regressão MAE*.

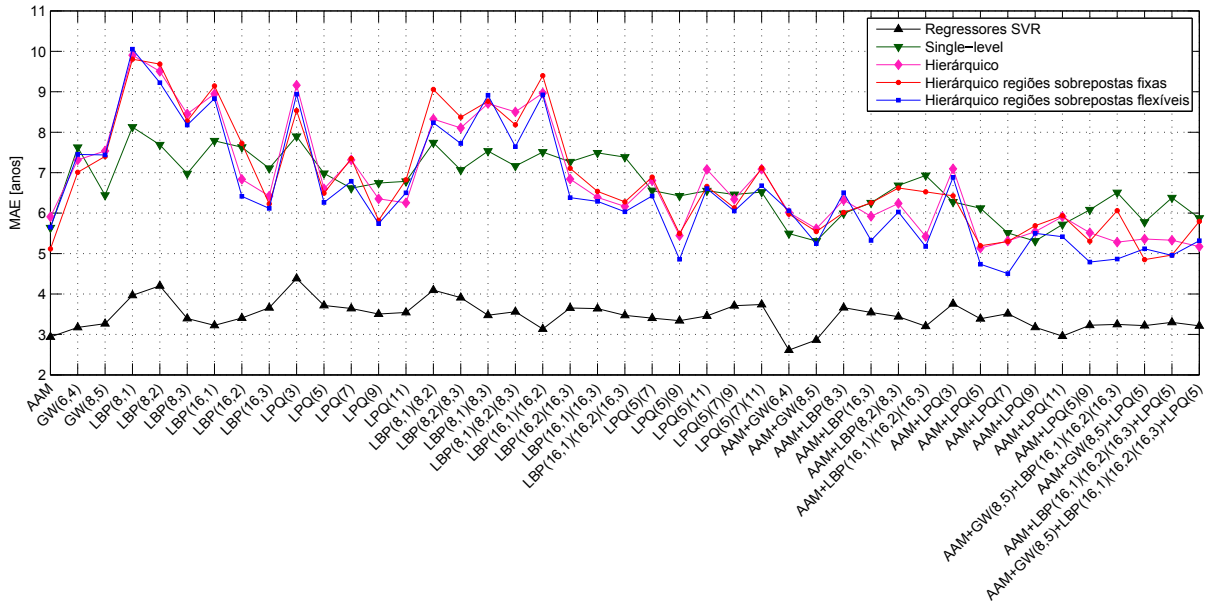


Figura 57: *MAE* de diferentes métodos de estimativa etária em diferentes conjuntos de características (FG-NET).

### 5.4.3 Estimativa hierárquica com regiões sobrepostas

Para compensar os erros gerados pela etapa de classificação é proposto neste trabalho a utilização de regiões sobrepostas fixas e flexíveis entre as faixas etárias na etapa da estimativa específica da idade. Para as regiões sobrepostas fixas cada faixa etária dos SVRs foi treinada com valores fixos de sobreposição, conforme o exemplo mostrado na Figura 58.

A Tabela 9 mostra alguns resultados em termos de *MAE* da estimativa hierárquica utilizando valores fixos para a sobreposição das faixas etárias. É possível perceber que os melhores resultados obtidos foram utilizando regiões sobrepostas com margens fixas de 3 e 5 anos para cada SVR. Porém, considerando o melhor *MAE* obtido com a estimativa hierárquica sem regiões sobrepostas fixas, que é 5,13 anos, se for utilizado uma região sobreposta com valores fixos de 5 anos o desempenho não melhorou, obtendo um *MAE* de 5,19 anos. Em geral, conforme resultados da Tabela 8, o desempenho utilizando regiões sobrepostas fixas de 5 anos entre as faixas etárias é similar a estimativa sem regiões

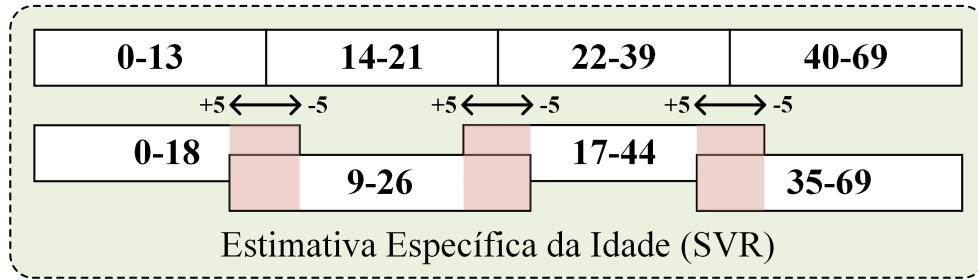


Figura 58: Exemplo dos regressores SVR da estimativa hierárquica com regiões sobrepostas fixas em 5 anos (FG-NET).

sobrepostas tendo um  $MAE$ , por exemplo, de 5,13 contra 5,19 anos, respectivamente, no caso da utilização das características  $AAM+LPQ_{5 \times 5}$ .

Tabela 9:  $MAE$  médio entre todos os conjuntos de características para algumas regiões sobrepostas fixas (FG-NET).

Idade sobreposta fixa	$MAE$ médio
$\pm 3$	5,67
$\pm 5$	5,79
$\pm 10$	6,18

Para tentar melhorar o desempenho dos regressores compensando os erros gerados na classificação, também foi avaliado a utilização de regiões sobrepostas flexíveis entre as faixas etárias. Foram analisados os erros gerados na etapa de classificação em faixas etárias para então decidir os valores iniciais das regiões flexíveis sobrepostas, a partir dos quais, experimentalmente, através do incremento ou decremento das idades das regiões sobrepostas foram escolhidos os valores que propiciaram os melhores resultados em termos de  $MAE$ . Na coluna *Hierárquico regiões sobrepostas fixas* da Tabela 8 são mostrados os resultados em diversos conjuntos de características da estimativa hierárquica utilizando regiões sobrepostas flexíveis. As idades para as regiões sobrepostas flexíveis entre os SVRs que obtiveram o melhor desempenho são mostrados na Figura 59 as quais foram utilizadas para os experimentos no conjunto de dados FG-NET. Por exemplo, a faixa etária utilizada para treinar o SVR 1 é de 0 a 28 anos, tendo uma região que sobrepõe até a faixa etária do SVR 3 treinado de 22 a 60 anos que sobrepõe parte do SVR 4 treinado de 40 a 69 anos. Essas regiões fazem com que os erros gerados na classificação sejam compensados nos SVRs treinados com faixas etárias mais abrangentes.

O desempenho em termos de  $MAE$  dos regressores SVR para a estimativa específica da idade com regiões sobrepostas flexíveis em diferentes conjuntos de características

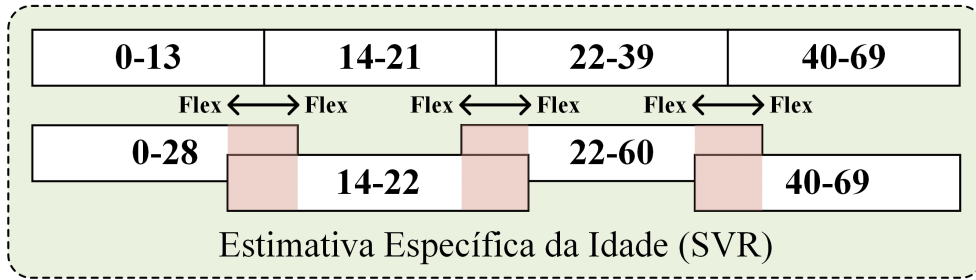


Figura 59: Idades utilizadas para as regiões sobrepostas flexíveis entre os SVRs (FG-NET).

é mostrado na Figura 60. Para o SVR 1, treinado na faixa etária de 0-28 anos o melhor desempenho isolado foi utilizando a combinação AAM+GW<sub>8,5</sub> com um *MAE* de 3,41 anos. O SVR 2 (14-22 anos) obteve o melhor *MAE*, 4,64 anos, com a combinação LBP<sub>(16,1)(16,2)(16,3)</sub>. Para a faixa etária de 22-60 anos (SVR 3) a combinação que obteve o melhor desempenho, *MAE* de 6,63 anos foi AAM+LPQ<sub>(5×5)(9×9)</sub>. Para o SVR 4 (40-69) a combinação AAM+LPQ<sub>7×7</sub> obteve o melhor desempenho, *MAE* de 0,09 anos. É possível perceber através do gráfico da Figura 60 que o SVR 1 teve o melhor desempenho geral, pois grande parte das imagens do conjunto de treinamento estão na faixa etária utilizada, já o SVR 4 tem um *MAE* que oscila muito entre os extremos, devido ao fato de ter sido utilizado poucos dados para o treinamento da faixa etária de 40 a 69 anos.

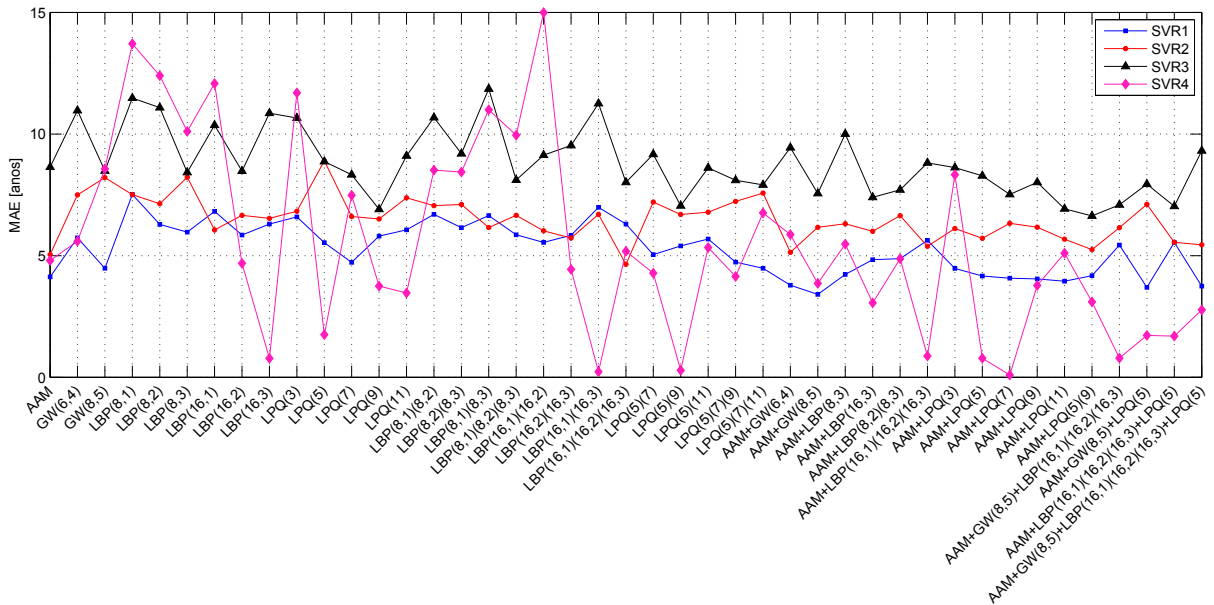


Figura 60: Desempenho em termos de *MAE* dos regressores SVR da estimativa hierárquica com regiões sobrepostas flexíveis em diferentes conjuntos de características (FG-NET).

De acordo com a Tabela 8, a combinação de características globais e locais AAM+LPQ<sub>7×7</sub> obteve o melhor desempenho, 4,50 anos de *MAE*. Utilizando margens fixas o desempenho a partir da mesma combinação de características foi de 5,30 anos de *MAE*, portanto houve um ganho considerável de 15,09% para a estimativa da idade. O desempenho médio para a estimativa hierárquica com regiões sobrepostas flexíveis foi de 6,53 anos de *MAE* considerando todos os conjuntos de características utilizados, sendo superior a todos os outros métodos utilizados nessa pesquisa, conforme é mostrado na Tabela 23 do Apêndice D com todos os resultados obtidos no conjunto FG-NET.

Os escores acumulativos (*CS*) utilizando níveis de erro entre 0 e 15 anos dos métodos de estimativa da idade utilizados nesta pesquisa para o conjunto FG-NET são mostrados na Figura 61. O método proposto utilizando a combinação de características holísticas, ou globais (AAM) e locais (LPQ<sub>7×7</sub>) obteve o melhor desempenho, pois é possível perceber que sua curva, em geral, está acima de todos os outros métodos. Ou seja, está mais próxima da curva ideal onde todos os níveis de erro teriam um *CS* de 100%, que significa que 100% das idades estimadas teriam um erro menor que o nível de erro (*MAE* em idade) definido. Por exemplo, no método proposto 70% das idades estimadas possuem um erro menor ou igual a 5 anos. Também, é possível perceber que com a combinação de características globais e locais o desempenho é superior em comparação a utilização de tais características isoladamente.

O desempenho da estimativa hierárquica com regiões sobrepostas flexíveis se mostrou superior ao método *single-level* e a estimativa hierárquica com margens fixas entre as classes. Devido as características relacionadas ao envelhecimento poderem ser diferentes de acordo com a faixa etária, a estimativa *single-level* apresentou limitações em estimar precisamente uma idade devido ter sido treinado em uma única faixa de idades. Portanto, o classificador hierárquico treinado em faixas menores obteve um melhor desempenho. Também foi avaliada a estimativa hierárquica da idade utilizando 7 classes, no entanto, os resultados não foram satisfatórios, pois os erros de classificação são maiores devido algumas classes não terem dados suficientes para treinamento e teste, conforme mostrados no Apêndice C.

A Tabela 8 mostra também as dimensões antes e depois da redução da dimensionalidade utilizando PCA para as imagens da base de dados FG-NET. A maior dimensão é derivada do extrator Gabor *wavelets* com 8 orientações e 5 escalas contendo 8.331 características para as 11 regiões de interesse. Devido a grande dimensionalidade das características extraídas é aconselhável utilizar algoritmos de redução da dimensão do espaço de

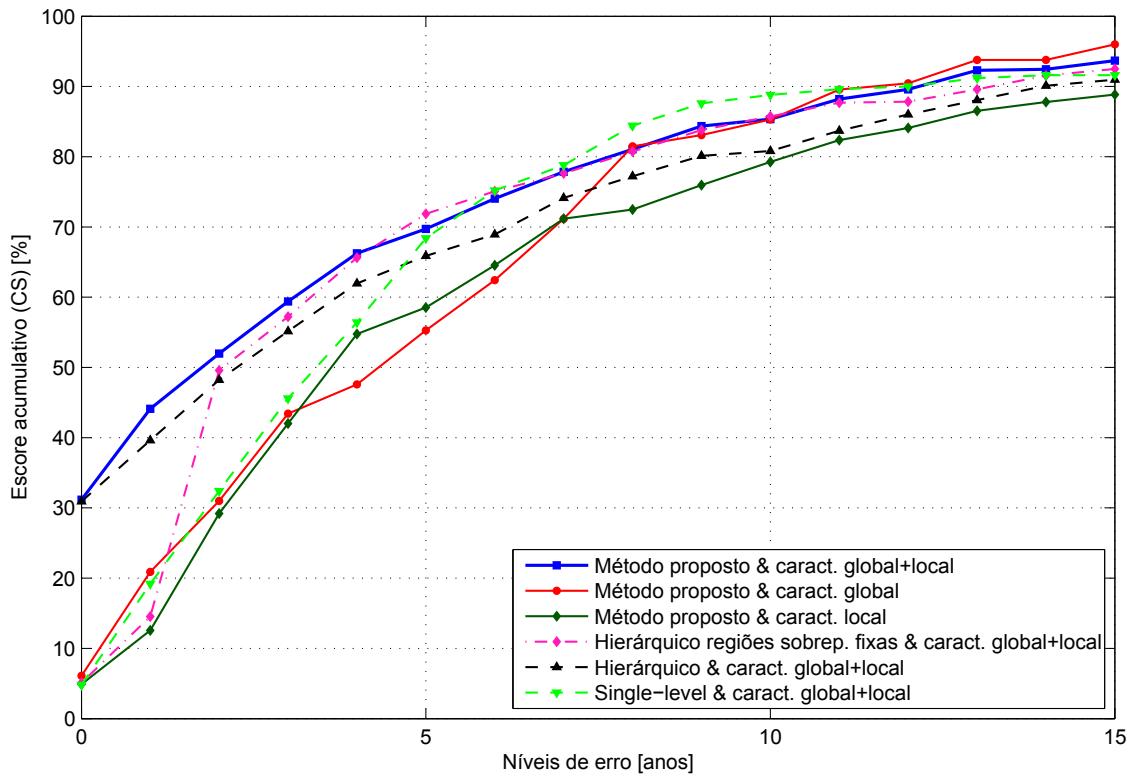


Figura 61: Escore acumulativo ( $CS$ ) dos métodos de estimativa da idade (FG-NET).

características, como por exemplo, o PCA utilizado neste trabalho. Com a aplicação do PCA nas características extraídas com  $GW_{8,5}$  apenas 180 características discriminantes das 8.331 foram selecionadas, reduzindo consideravelmente o custo computacional. As novas dimensões das características utilizando PCA foram decididas experimentalmente através do desempenho da estimativa hierárquica em cada conjunto de características. O desempenho é medido a partir do incremento da dimensão PCA (número de componentes principais) e a dimensão com o melhor desempenho é selecionado.

Uma comparação em termos de dimensão e  $MAE$  (idade) entre a concatenação de algumas características e a aplicação do método de fusão de características proposto é mostrada na Tabela 10. É possível perceber que o método de fusão das características tem um desempenho melhor do que apenas concatená-las após a redução da dimensionalidade. Na método de fusão proposto as características são primeiramente normalizadas pelo método z-escore, concatenadas e então aplicado PCA novamente. A fusão das características globais (AAM) e locais ( $LPQ_{5 \times 5}$ ), por exemplo, utilizando apenas a concatenação teve um  $MAE$  de 6,46 e com uma dimensão de 262, enquanto aplicando o método de fusão o  $MAE$  é reduzido consideravelmente para 5,13, além da dimensão ser reduzida também para 215 características discriminantes.



Tabela 10: Comparação em termos de dimensão e  $MAE$  (idade) utilizando a concatenação das características e a aplicação do método de fusão de características (FG-NET).

Características	Concatenação $MAE$ /Dimensão	Método de fusão $MAE$ /Dimensão
AAM+LPQ <sub>3×3</sub>	7,48/194	7,10/170
AAM+LPQ <sub>5×5</sub>	6,46/262	5,13/215
AAM+LPQ <sub>7×7</sub>	6,00/259	5,31/158
AAM+LPQ <sub>9×9</sub>	5,76/257	5,54/129
AAM+LPQ <sub>11×11</sub>	6,00/254	5,91/102

Nesta seção foram apresentados os resultados dos experimentos realizados sob o conjunto de dados FG-NET para avaliar o desempenho do sistema proposto e também de outros métodos para efeitos de comparação. O método proposto, estimativa hierárquica da idade com regiões sobrepostas flexíveis, avaliado no conjunto de dados FG-NET teve um desempenho médio de 6,53 anos de  $MAE$ , sendo superior a todos os outros métodos utilizados nessa pesquisa como a estimativa *single-level*, estimativa hierárquica com margens fixas e estimativa hierárquica com regiões sobrepostas fixas. O melhor desempenho em termos de  $MAE$ , 4,50 anos, foi atingido utilizando a combinação de características globais e locais AAM+LPQ<sub>7×7</sub>. As características locais extraídas utilizando o descritor LPQ foram as mais discriminantes mostrando-se muito robusto não apenas ao borramento mas também a outros desafios, como variações na iluminação e expressão facial presentes no conjunto de dados utilizado, superando outros métodos utilizados em outros trabalhos como Gabor *wavelets* e LBP. Na próxima seção serão apresentados os resultados dos experimentos para avaliar o desempenho do sistema proposto sob o conjunto de dados MORPH.

#### 5.4.4 Experimentos utilizando o Protocolo LOPO

Conforme descrito na Seção 5.3, o método proposto foi também avaliado no conjunto FG-NET *Aging* através do protocolo LOPO. A comparação dos resultados da estimativa hierárquica com margens flexíveis utilizando o protocolo dos experimentos anteriores (75% para treinamento e 25% para teste) e o LOPO é mostrada na Tabela 11. O desempenho utilizando a combinação de características globais e locais AAM+LPQ<sub>7×7</sub> foi de 4,78 anos, um pouco inferior quando comparado ao protocolo 75%/25%, que obteve 4,50, porém, esse esquema garante que uma pessoa não esteja no conjunto de teste e treinamento ao mesmo tempo eliminando assim a dependência dos dados nos resultados do experimento.

Os escores acumulativos utilizando níveis de erro entre 0 e 15 anos da estimativa

Tabela 11: Comparação dos resultados da estimativa hierárquica com margens flexíveis utilizando o protocolo dos experimentos anteriores (75% para treinamento e 25% para teste) e o LOPO (FG-NET).

Características	75%/25% ( <i>MAE</i> )	LOPO ( <i>MAE</i> )
AAM	5,65	5,19
LPQ <sub>7×7</sub>	6,79	5,95
<b>AAM+LPQ<sub>7×7</sub></b>	<b>4,50</b>	<b>4,78</b>

hierárquica com margens flexíveis com a combinação de características globais (AAM) e locais (LPQ<sub>7×7</sub>) são mostrados na Figura 62. O método proposto utilizando o protocolo LOPO possui uma curva similar a utilizar 75% dos dados para treinamento e o restante para teste. Sendo levemente superior no nível de erro 5, por exemplo, possuindo aproximadamente 71% dos dados estimados com um *MAE* menor ou igual a 5 anos.

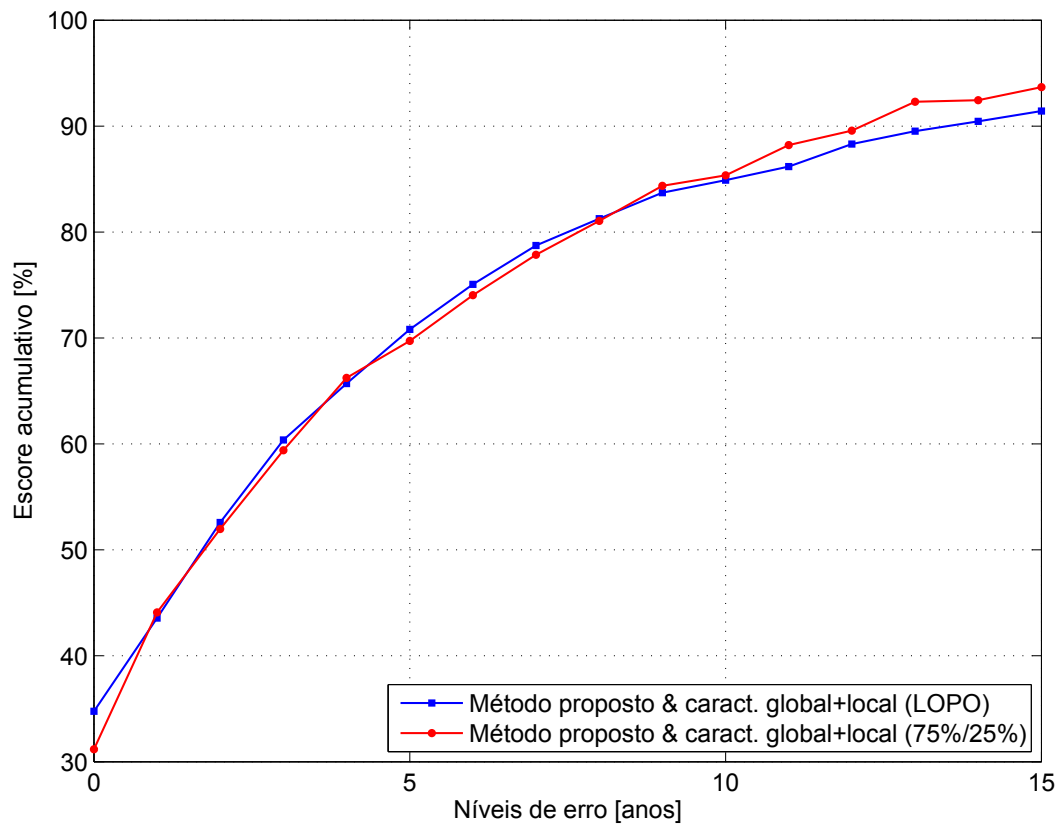


Figura 62: Escore acumulativo (*CS*) do método proposto utilizando os protocolos 75%/25% e LOPO (FG-NET).

## 5.5 Conjunto de dados MORPH *Album 2*

Nesta seção são apresentados os resultados da estimativa da idade no conjunto de dados MORPH *Album 2* utilizando métodos de estimação hierárquica com margens fixas e com regiões sobrepostas. Foram selecionados apenas as características e combinações que levaram aos melhores resultados no conjunto de dados FG-NET.

### 5.5.1 Estimativa hierárquica com margens fixas

#### 5.5.1.1 Classificação em faixas etárias

As faixas etárias de cada classe foram definidas de acordo com a distribuição dos dados do conjunto MORPH *Album 2* para que cada classe tenha dados suficientes para treinamento e teste de uma forma bem distribuída. A Tabela 12 mostra a distribuição do conjunto MORPH *Album 2* dividida em faixas etárias de 10 anos (décadas) sendo possível observar que 84% dos dados estão entre 16 e 45 anos. As faixas etárias e a distribuição dos dados de treinamento e teste para cada classe são mostradas na Tabela 13.

Tabela 12: Distribuição do conjunto de dados em décadas (MORPH).

Faixa etária	Número de amostras
16-25	1.420
26-35	1.486
36-45	1.726
46-55	711
56-65	120
66-77	11
Total	5.474

Tabela 13: Faixas etárias e distribuição dos dados de treinamento e teste para cada classe (MORPH).

Classe	Faixa etária	Treinamento	Teste
1	16-25	1.065	520
2	26-35	1.115	369
3	36-45	1.295	266
4	46-77	632	212
Total	16-77	4.107	1.367

A Figura 63 mostra o desempenho da classificação em cada conjunto de características utilizando kernel linear. A maior taxa de acerto foi obtida com a característica

local  $LPQ_{9 \times 9}$  tendo uma taxa de 51,06% de instâncias classificadas corretamente. Nota-se no gráfico da Figura 63 que a combinação de características globais e locais não são tão eficientes para esse conjunto de dados, pois a faixa etária é de 17 a 77 anos e o AAM é mais eficiente para a descrição do formato facial, sendo mais proeminentes na faixa etária de 0 a 18 anos, portanto as características globais não foram muito relevantes no conjunto de dados MORPH.

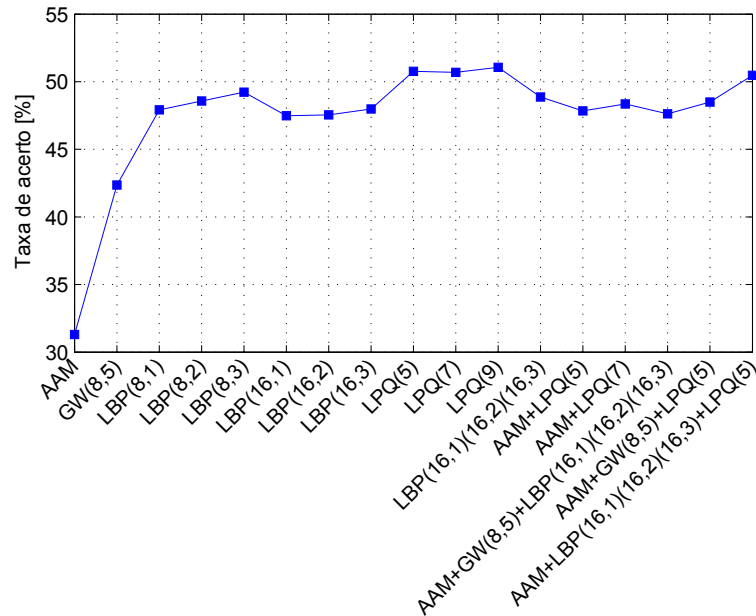


Figura 63: Desempenho da classificação utilizando kernel linear (LIBLINEAR) em diferentes conjuntos de características (MORPH).

A matriz de confusão da classificação em faixas etárias utilizando características locais  $LPQ_{(7 \times 7)}$  as quais obtiveram o melhor desempenho da estimativa hierárquica da idade com margens fixas (Tabela 24) é mostrada na Figura 64. A taxa de acerto geral da classificação foi de 50,7%. A classe 1 (16-25 anos) foi a que obteve a melhor taxa de acerto na classificação, 80,3%, tendo um erro de 19,7% sendo que parte das instâncias de teste foram classificadas dentro das classes 2 (26-35 anos) e 3 (36-45 anos). A classe 2 obteve um erro de 67,1% devido a classificar erroneamente grande parte das instâncias de teste na classe 1. A classe 3 teve um erro de 63,6%, classificando erroneamente algumas instâncias de teste nas classes 1, 2 e 4. Isso se deve ao fato de existirem indivíduos das classe 3 que podem ser classificados dentro das classes 1, 2 ou 4 por possuírem características semelhantes como poucas rugas e sem marcas na pele devido a fatores genéticos, estilos de vida, saúde etc. A classe 4 (46-77 anos) teve uma taxa de acerto de 61,4%, classificando erroneamente algumas instâncias nas classes 1, 2 e 3 pelos mesmos motivos descritos anteriormente. A matriz de confusão mostra claramente a tendência da base MORPH

em ser mais precisa na faixa etária de 16 a 25 anos e também na faixa etária de 46 a 77 anos. As confusões ocorreram também devido ao não tratamento das margens duras (*hard boundaries*) do classificador, onde um indivíduo de 26 anos (classe 2), por exemplo, pode ser facilmente classificado dentro da classe 1 que possui uma faixa etária de 16 a 25 anos.

		Classe verdadeira				Taxa de acerto	
		1	2	3	4		
Classe predita	1	285 20.8%	52 3.8%	18 1.3%	0 0.0%	80.3%	19.7%
	2	166 12.1%	122 8.9%	74 5.4%	9 0.7%	32.9%	67.1%
	3	68 5.0%	114 8.3%	157 11.5%	92 6.7%	36.4%	63.6%
	4	12 0.9%	22 1.6%	47 3.4%	129 9.4%	61.4%	38.6%

Figura 64: Matriz de confusão da classificação em faixas etárias utilizando características locais LPQ<sub>(7×7)</sub> (MORPH).

### 5.5.1.2 Estimativa específica da idade

Para a etapa da estimativa específica da idade foram utilizadas as mesmas faixas etárias e distribuições de dados descritas anteriormente na etapa da classificação (Tabela 13) para definir os regressores SVR. A característica que obteve o melhor desempenho para a estimativa hierárquica com margens fixas foi a característica local LPQ<sub>(5×5)</sub> com um *MAE* de 6,67 anos. Considerando o desempenho individual de cada regressor, o SVR 1, treinado na faixa etária de 16-25 anos obteve um *MAE* de 6,90 anos, o SVR 2 (26-35 anos) 7,44 anos, o SVR 3 (36-45 anos) 5,95 anos e o SVR 4 (46-77 anos) obteve um *MAE* de 6,41 anos. Na Figura 65 são mostrados os resultados da estimativa hierárquica (linha *Hierárquico*) em diferentes conjuntos de características. Na Tabela 14 são mostrados os resultados numéricos de alguns conjuntos de características utilizados para a estimativa hierárquica da idade (coluna *Hierárquico MAE*).

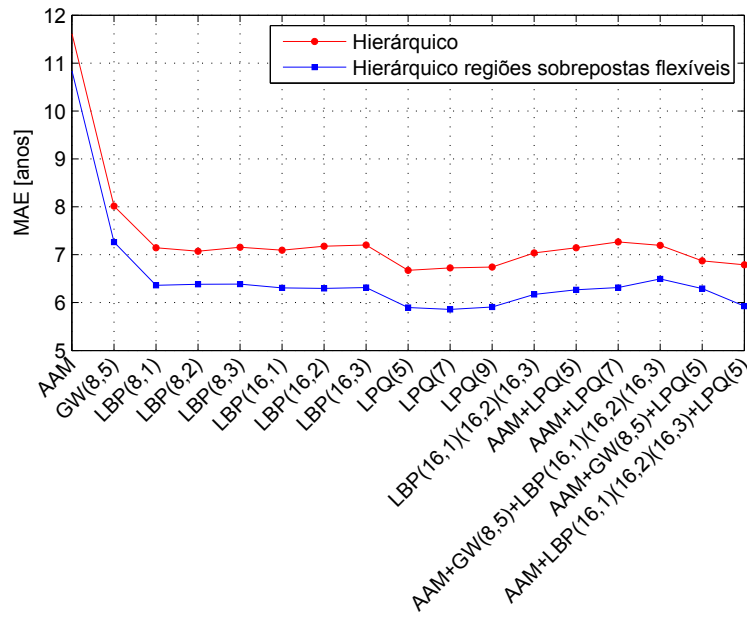


Figura 65: Comparação em termos de  $MAE$  entre a estimativa hierárquica da idade com margens fixas e com regiões sobrepostas flexíveis em diferentes conjuntos de características (MORPH).

## 5.5.2 Estimativa hierárquica com regiões sobrepostas

Para os experimentos realizados no conjunto de dados MORPH foram utilizadas regiões sobrepostas flexíveis apenas, pois o método obteve melhores resultados no conjunto de dados FG-NET. Os mesmos procedimentos para a definição das regiões sobrepostas flexíveis no conjunto FG-NET foram realizados no conjunto MORPH. Na coluna *Hierárquico regiões sobrepostas fixas* da Tabela 14 são mostrados os resultados em diversos conjuntos de características da estimativa hierárquica utilizando regiões sobrepostas flexíveis. As idades para as regiões sobrepostas flexíveis entre os SVRs que propiciaram o melhor desempenho são mostrados na Figura 66, as quais foram utilizados para os experimentos no conjunto de dados MORPH. Por exemplo, a faixa etária utilizada para treinar o SVR 1 é de 16 a 35 anos, tendo uma região que sobrepõe até a faixa etária do SVR 2, treinado de 16 a 45 anos que sobrepõe parte do SVR 1 e 3, treinado de 26 a 62 anos sobrepondo regiões dos SVRs 2 e 4, e o SVR 4 é treinado de 37 a 77 anos, o qual sobrepõe uma faixa do SVR 3. Essas regiões fazem com que os erros gerados na classificação sejam compensados nos SVRs treinados com faixas etárias mais abrangentes.

O desempenho médio obtido para a estimativa hierárquica com regiões sobrepostas flexíveis foi de 6,55 anos de  $MAE$ , sendo superior ao método com margens fixas que obteve um  $MAE$  médio de 7,35 anos. O melhor desempenho em termos de  $MAE$ , 5,86

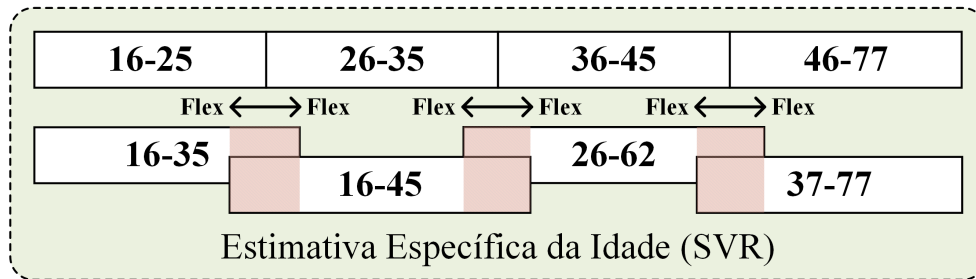


Figura 66: Idades utilizadas para as regiões sobrepostas flexíveis entre os SVRs (MORPH).

Tabela 14: Resultados de diferentes métodos de estimação etária em diferentes conjuntos de características (MORPH).

Características	Classificação Taxa de acerto [%]	Hierárquico <i>MAE</i>	Hierárquico regiões sobrepostas flexíveis <i>MAE</i>
AAM	31,31	11,61	10,84
GW <sub>8,5</sub>	42,36	8,02	7,26
LBP <sub>16,2</sub>	47,55	7,18	6,29
LPQ <sub>7×7</sub>	50,69	6,72	<b>5,86</b>
LBP <sub>(16,1)(16,2)(16,3)</sub>	48,87	7,04	6,17
AAM+LPQ <sub>7×7</sub>	48,35	7,27	6,31
AAM+GW <sub>8,5</sub> +LBP <sub>(16,1)(16,2)(16,3)</sub>	47,62	7,19	6,49
AAM+GW <sub>8,5</sub> +LPQ <sub>5×5</sub>	48,50	6,87	6,29
AAM+LBP <sub>(16,1)(16,2)(16,3)</sub> +LPQ <sub>5×5</sub>	50,48	6,79	5,93

anos, utilizando as características locais LPQ<sub>7×7</sub> foi obtido a partir da técnica de sobreposição flexível das faixas etárias na etapa da estimativa específica da idade. Utilizando margens fixas o desempenho utilizando a mesma combinação de características foi de 6,72 anos de *MAE*, portanto houve um ganho considerável de 14,68% para a estimativa da idade, sendo aproximado ao ganho de 15,09% obtido no conjunto FG-NET mostrando a robustez do método proposto. Também, é possível concluir que as características locais extraídas utilizando o descritor LPQ foram as mais discriminantes em ambos os conjuntos de dados, MORPH e FG-NET, mostrando os melhores resultados em termos de *MAE*. No Apêndice E são mostrados todos os resultados obtidos no conjunto MORPH.

O desempenho em termos de *MAE* dos regressores SVR para a estimativa específica da idade com regiões sobrepostas flexíveis em diferentes conjuntos de características é mostrado na Figura 67. Para o SVR 1, treinado na faixa etária de 16-45 anos, o melhor

desempenho isolado foi utilizando a combinação  $AAM+LBP_{(16,1)(16,2)(16,3)}+LPQ_{5 \times 5}$  com um  $MAE$  de 5,88 anos. O SVR 2 (14-22 anos) obteve o melhor  $MAE$ , 6,19 anos, com a combinação  $LPQ_{9 \times 9}$ . Para a faixa etária de 26-62 anos (SVR 3) a combinação que obteve o melhor desempenho,  $MAE$  de 5,28 anos, foi o  $LPQ_{5 \times 5}$ . Para o SVR 4 (37-77) a combinação  $AAM+GW_{8,5}+LPQ_{5 \times 5}$  obteve o melhor desempenho,  $MAE$  de 5,09 anos. É possível perceber através do gráfico da Figura 67 que o SVR 4 teve o melhor desempenho geral seguido do SVR 3, SVR 1 e SVR 2.

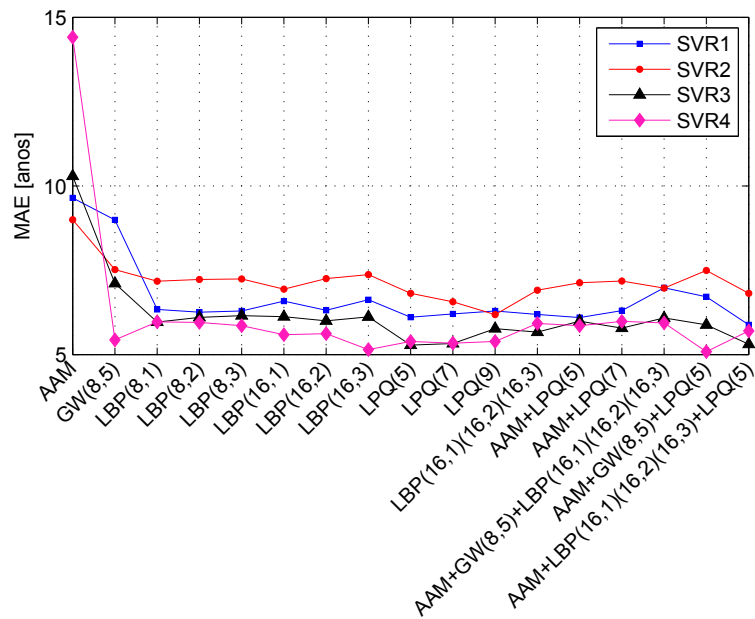


Figura 67: Desempenho em termos de  $MAE$  dos regressores SVR da estimativa hierárquica com regiões sobrepostas flexíveis em diferentes conjuntos de características (MORPH).

Os escores acumulativos ( $CS$ ) utilizando níveis de erro entre 0 e 15 anos dos métodos de estimativa da idade utilizados nesta pesquisa para o conjunto MORPH são mostrados na Figura 68. O método proposto utilizando a combinação de características holísticas, ou globais (AAM) e locais ( $LPQ_{7 \times 7}$ ) obteve um desempenho levemente inferior a utilização de apenas características locais, principalmente nos níveis de erro entre 5 e 10 anos. Esse desempenho inferior ocorre devido a faixa etária do conjunto MORPH ser de 17 a 77 anos e o AAM ser mais eficiente para a descrição do formato facial, sendo mais proeminentes na faixa etária de 0 a 18 anos, portanto tais características globais não foram muito relevantes no conjunto de dados MORPH como é possível perceber no gráfico (curva *Método proposto* e *caract. global+local*).

Nesta seção foram apresentados os resultados dos experimentos realizados sob o conjunto de dados MORPH para avaliar o desempenho do sistema proposto e também



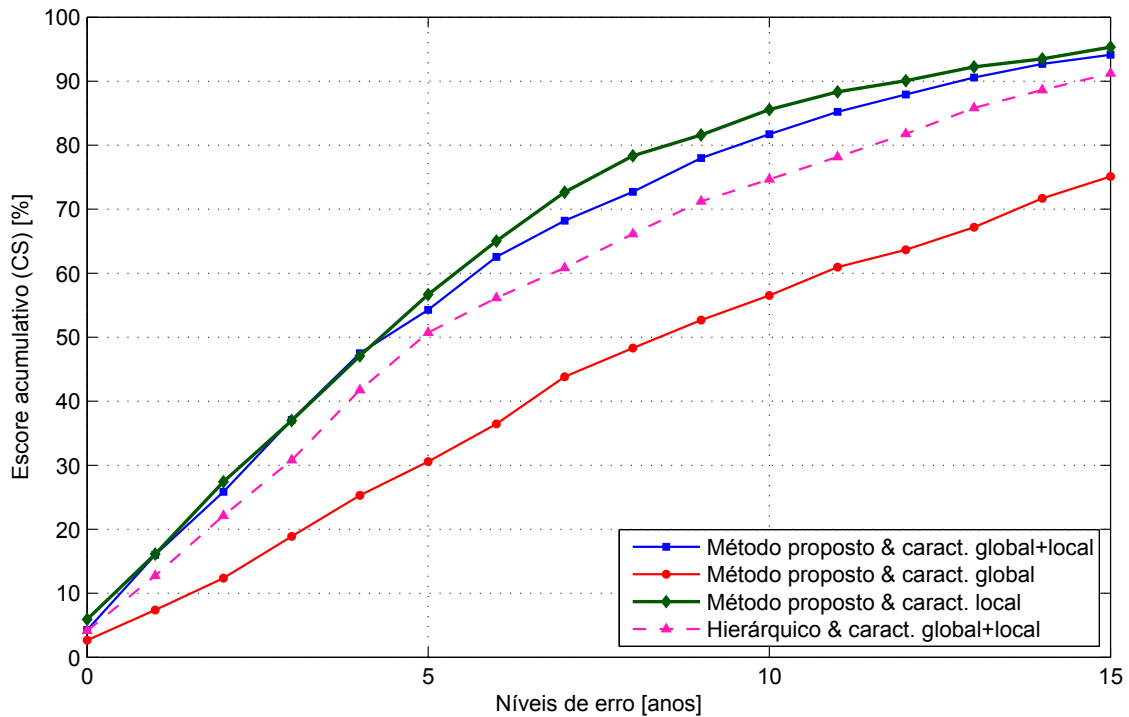


Figura 68: Escore acumulativo ( $CS$ ) dos métodos de estimativa da idade (MORPH).

de outros métodos para efeitos de comparação. O método proposto, estimativa hierárquica da idade com regiões sobrepostas flexíveis, avaliado no conjunto de dados MORPH teve um desempenho médio de 6,55 anos de  $MAE$ , sendo superior aos outros métodos utilizados como a estimativa hierárquica com margens fixas. O melhor desempenho em termos de  $MAE$ , 5,86 anos, foi obtido utilizando as características locais  $LPQ_{7 \times 7}$ . As características locais extraídas utilizando o descritor  $LPQ$  foram, mais uma vez, as mais discriminantes quando comparadas a características extraídas a partir de técnicas tradicionais como Gabor *wavelets* e LBP. Na próxima seção serão apresentados os resultados dos experimentos para avaliar o desempenho do sistema proposto sob a combinação dos conjuntos de dados FG-NET e MORPH.

## 5.6 Combinação dos conjuntos de dados FG-NET e MORPH

Os conjuntos de dados FG-NET e MORPH foram combinados para ter uma maior distribuição de dados em determinadas faixas etárias, como é o caso da faixa etária entre 0 e 16 anos que o conjunto MORPH não possui nenhuma imagem, enquanto grande parte das imagens do conjunto FG-NET estão nessa faixa etária. Assim como imagens entre 70 e 77

anos que apenas o conjunto MORPH possui. Foram selecionados apenas as características e combinações que obtiveram os melhores resultados nos conjuntos de dados FG-NET e MORPH.

## 5.6.1 Estimativa hierárquica com margens fixas

### 5.6.1.1 Classificação em faixas etárias

As faixas etárias de cada classe foram definidas de acordo com a distribuição dos dados do conjunto FG-NET+MORPH mostrada na Figura 69 para que cada classe tenha dados suficientes para treinamento e teste de uma forma bem distribuída. Também, as classes foram definidas de acordo com os grupos etários: crianças, jovens, adultos e idosos. As faixas etárias e a distribuição dos dados de treinamento e teste para cada classe são mostradas na Tabela 15.

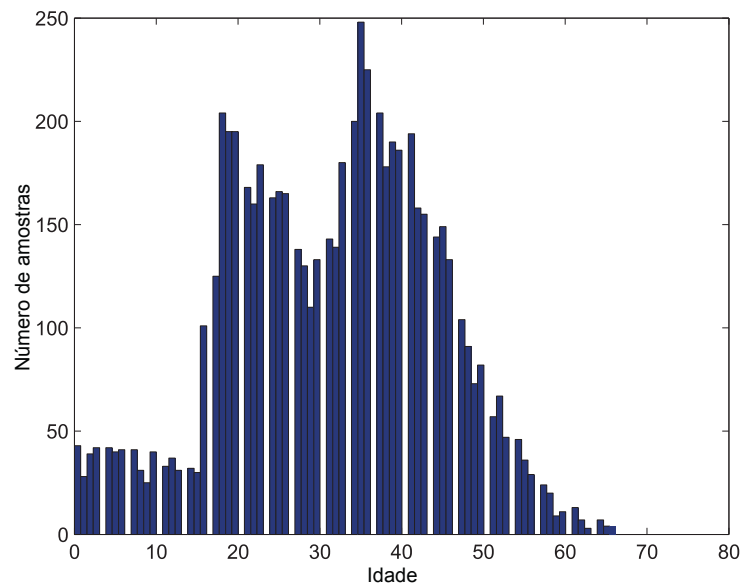


Figura 69: Distribuição de idade após a combinação dos conjuntos de dados FG-NET e MORPH.

A Figura 70 mostra o desempenho da classificação em cada conjunto de características utilizado. A maior taxa de acerto foi atingida com a combinação de características globais e locais AAM+LPQ<sub>7×7</sub> tendo uma taxa de 64,46% de instâncias classificadas corretamente.

A matriz de confusão da classificação em faixas etárias utilizando a combinação de características globais e locais AAM+LPQ<sub>(7×7)</sub> as quais obtiveram o melhor desempenho da estimativa hierárquica da idade com margens fixas (Tabela 16) é mostrada na Figura 71.

Tabela 15: Distribuição dos dados de treinamento e teste para cada regressor SVR da estimativa específica da idade (FG-NET+MORPH).

Classe	Faixa etária	Treinamento	Teste
1	0-13	385	113
2	14-21	788	95
3	22-39	2.288	1.139
4	40-77	1.397	271
Total	0-77	4.858	1.618

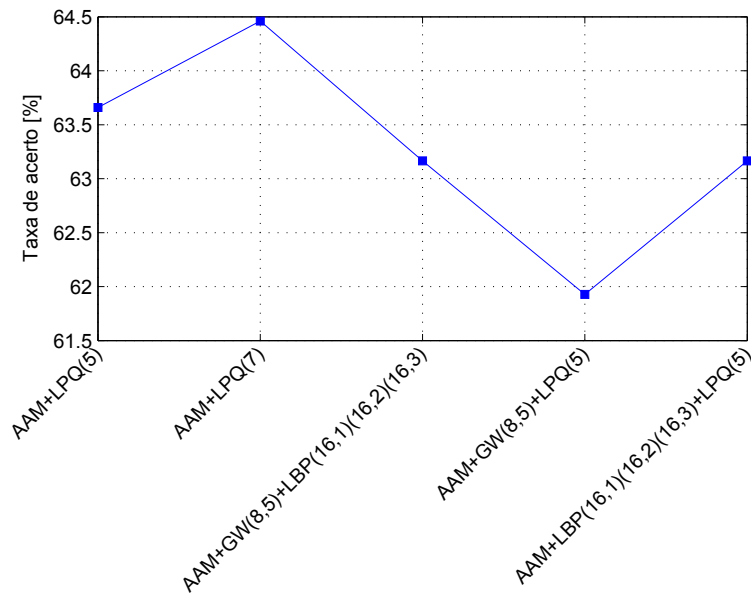


Figura 70: Desempenho da classificação em diferentes conjuntos de características (FG-NET+MORPH).

A taxa de acerto geral da classificação foi de 64,5%. As classes 1 (0-13 anos) e 3 (22-39 anos) foram as que levaram as melhores precisões na classificação, 88,3% e 88,2%, respectivamente. A classe 2 (14-21 anos) obteve um erro de 83,6% devido a classificar erroneamente a maioria das instâncias de teste na classe 3. Isso se deve ao fato de existirem indivíduos da classe 2 que podem ser facilmente classificados dentro da classe 3 por possuírem características semelhantes. A classe 4 (40-77 anos) obteve um erro de 54% devido a classificar erroneamente a maioria das instâncias de teste na classe 3. A matriz de confusão mostra claramente que a combinação dos conjuntos de dados FG-NET e MORPH, de acordo com a nova distribuição dos dados, tende a ser mais precisa nas faixa etárias de 0 a 13 anos (Classe 1) e 22 e 39 anos (classe 3). Isso é devido a Classe 1 possuir características mais discriminantes relacionadas ao formato facial extraídas pelo AAM e a Classe 3 possuir mais imagens faciais para o treinamento.

		Taxa de acerto				
Classe predita	1	113 7.0%	10 0.6%	3 0.2%	2 0.1%	88.3% 11.7%
	2	0 0.0%	43 2.7%	216 13.3%	3 0.2%	16.4% 83.6%
	3	0 0.0%	38 2.3%	673 41.6%	52 3.2%	88.2% 11.8%
	4	0 0.0%	4 0.2%	247 15.3%	214 13.2%	46.0% 54.0%
		1	2	3	4	
		Classe verdadeira				

Figura 71: Matriz de confusão da classificação em faixas etárias utilizando a combinação de características AAM e  $LPQ_{7 \times 7}$  (FG-NET+MORPH).

### 5.6.1.2 Estimativa específica da idade

Para a etapa da estimativa específica da idade foram utilizadas as mesmas faixas etárias e distribuições de dados descritas anteriormente na etapa da classificação (Tabela 15) para definir os regressores SVR. A característica que obteve o melhor desempenho para a estimativa hierárquica com margens fixas foi a combinação de características globais e locais AAM+ $LPQ_{(7 \times 7)}$  com um  $MAE$  de 5,55 anos. Considerando o desempenho individual de cada regressor, o SVR 1, treinado na faixa etária de 0-13 anos obteve um  $MAE$  de 2,11 anos, o SVR 2 (14-21 anos) 5,51 anos, o SVR 3 (22-39 anos) 7,70 anos e o SVR 4 (40-77 anos) obteve um  $MAE$  de 6,90 anos. Na Figura 72 são mostrados os resultados da estimativa hierárquica (linha *Hierárquico*) em diferentes conjuntos de características. Na Tabela 16 são mostrados os resultados numéricos de alguns conjuntos de características utilizados para a estimativa hierárquica da idade (coluna *Hierárquico MAE*).

### 5.6.2 Estimativa hierárquica com regiões sobrepostas

Para os experimentos realizados no conjunto de dados FG-NET+MORPH foram utilizadas regiões sobrepostas flexíveis apenas, pois o método obteve melhores resultados no conjunto de dados FG-NET e MORPH. Os mesmos procedimentos para a definição das regiões sobrepostas flexíveis no conjunto FG-NET e MORPH foram realizados no con-

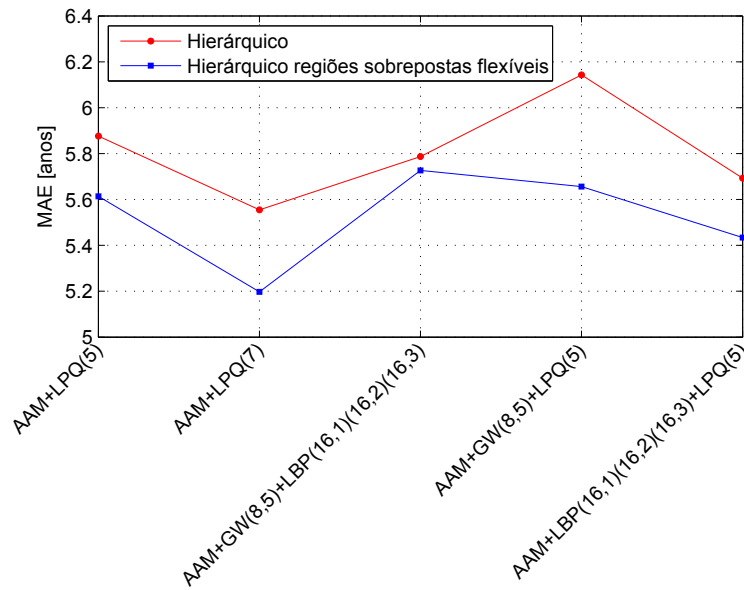


Figura 72: Comparação em termos de  $MAE$  entre a estimativa hierárquica da idade com margens fixas e com regiões sobrepostas flexíveis em diferentes conjuntos de características (FG-NET+MORPH).

Tabela 16: Resultados da estimativa hierárquica com margens fixas e flexíveis em diferentes conjuntos de características, mostrando também a dimensão das características após a redução da dimensionalidade utilizando PCA (FG-NET+MORPH).

Características	Classificação Taxa de acerto [%]	Hierárquico $MAE$	Hierárquico regiões sobrepostas flexíveis $MAE$
AAM+LPQ <sub>5×5</sub>	63,66	5,88	5,61
AAM+LPQ <sub>7×7</sub>	<b>64,46</b>	<b>5,55</b>	<b>5,20</b>
AAM+GW <sub>8,5</sub> +LBP <sub>(16,1)(16,2)(16,3)</sub>	63,16	5,79	5,73
AAM+GW <sub>8,5</sub> +LPQ <sub>5×5</sub>	61,93	6,14	5,66
AAM+LBP <sub>(16,1)(16,2)(16,3)</sub> +LPQ <sub>5×5</sub>	63,16	5,69	5,43
<b>Média geral</b>	63,28	5,81	5,53

junto FG-NET+MORPH. Na coluna *Hierárquico regiões sobrepostas fixas* da Tabela 16 são mostrados os resultados em diversos conjuntos de características da estimativa hierárquica utilizando regiões sobrepostas flexíveis. As idades para as regiões sobrepostas flexíveis entre os SVRs que obtiveram o melhor desempenho são mostrados na Figura 73 as quais foram utilizados para os experimentos no conjunto de dados FGNET+MORPH. Por exemplo, a faixa etária utilizada para treinar o SVR 1 é de 0 a 15 anos, tendo uma região que sobrepõe até a faixa etária do SVR 2, treinado de 8 a 25 anos, sobrepondo

parte do SVR 1 e 3, treinado de 19 a 66 anos, e o SVR 4 é treinado de 20 a 77 anos, sobrepondo uma região até o SVR 2. Essas regiões fazem com que os erros gerados na classificação sejam compensados nos SVRs treinados com faixas etárias mais abrangentes.

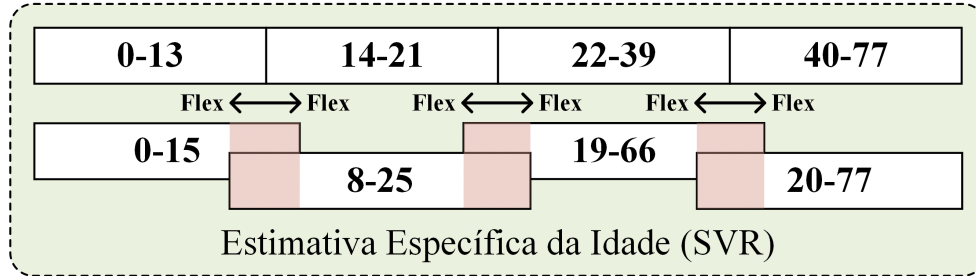


Figura 73: Idades utilizadas para as regiões sobrepostas flexíveis entre os SVRs (FG-NET+MORPH).

Conforme mostram os resultados da Tabela 16 para a estimativa hierárquica com regiões sobrepostas flexíveis o desempenho médio obtido foi de 5,53 anos de  $MAE$ , sendo superior ao método com margens fixas que obteve um  $MAE$  médio de 5,81 anos. O melhor desempenho em termos de  $MAE$ , 5,20 anos, utilizando a combinação de características globais e locais AAM+LPQ $_{7 \times 7}$  foi obtido a partir da técnica de sobreposição flexível das faixas etárias na etapa da estimativa específica da idade. Utilizando margens fixas o desempenho utilizando a mesma combinação de características foi de 5,55 anos de  $MAE$ , portanto houve um ganho de 6,73% para a estimativa da idade.

O desempenho em termos de  $MAE$  dos regressores SVR para a estimativa específica da idade com regiões sobrepostas flexíveis em diferentes conjuntos de características é mostrado na Figura 74. Para o SVR 1, treinado na faixa etária de 0-15 anos, o melhor desempenho isolado foi utilizando a combinação AAM+LPQ $_{(7 \times 7)}$  com um  $MAE$  de 1,97 anos. O SVR 2 (8-25 anos) obteve o melhor  $MAE$ , 4,51 anos, com a combinação AAM+GW $_{8,5}$ +LBP $_{(16,1)(16,2)(16,3)}$ . Para a faixa etária de 19-66 anos (SVR 3) a combinação que obteve o melhor desempenho,  $MAE$  de 6,35 anos, foi AAM+LBP $_{(16,1)(16,2)(16,3)}$ +LPQ $_{5 \times 5}$ . Para o SVR 4 (20-77) a combinação AAM+LPQ $_{(7 \times 7)}$  obteve o melhor desempenho,  $MAE$  de 6,63 anos. É possível perceber através do gráfico da Figura 74 que o SVR 1 teve o melhor desempenho geral mostrando mais uma vez que nessa faixa etária o erro da estimativa da idade é menor devido possuir características mais discriminantes com relação ao formato facial e textura da pele.

Os escores acumulativos ( $CS$ ) utilizando níveis de erro entre 0 e 15 anos da estimativa hierárquica com regiões sobrepostas flexíveis para o conjunto FG-NET+MORPH é mostrado na Figura 75. O método proposto, utilizando a combinação de características

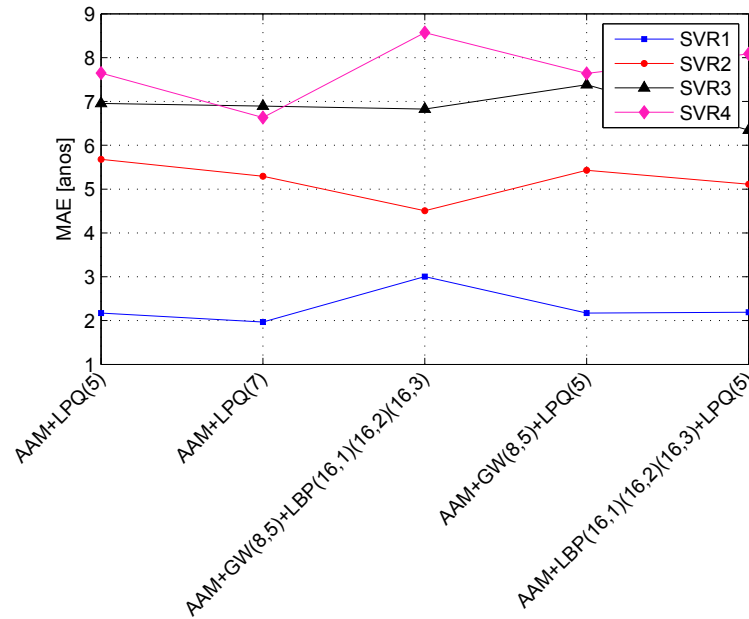


Figura 74: Desempenho em termos de  $MAE$  dos regressores SVR da estimativa hierárquica com regiões sobrepostas flexíveis em diferentes conjuntos de características (FG-NET+MORPH).

holísticas, ou globais (AAM) e locais ( $LPQ_{7 \times 7}$ ) obteve o melhor desempenho. É possível perceber que mais de 85% das idades são estimadas com um erro menor que 10 anos.

Nesta seção foram apresentados os resultados dos experimentos realizados sob a combinação dos conjuntos de dados FG-NET e MORPH para avaliar o desempenho e robustez do sistema proposto através de um conjunto de dados maior e com uma faixa etária mais abrangente. A estimativa hierárquica da idade com regiões sobrepostas flexíveis foi a que obteve o melhor desempenho confirmando a robustez do sistema proposto obtendo um desempenho médio de 5,53 anos de  $MAE$ . O melhor desempenho em termos de  $MAE$ , 5,20 anos, foi utilizando a combinação de características globais e locais AAM+ $LPQ_{7 \times 7}$ , que mostra a eficiência do método proposto ao utilizar a combinação de características globais (AAM) e locais através de características de fase local extraídas a partir do operador LPQ. Na próxima seção serão apresentados os resultados da estimativa realizada por seres humanos, assim como comparações com o método computacional proposto.

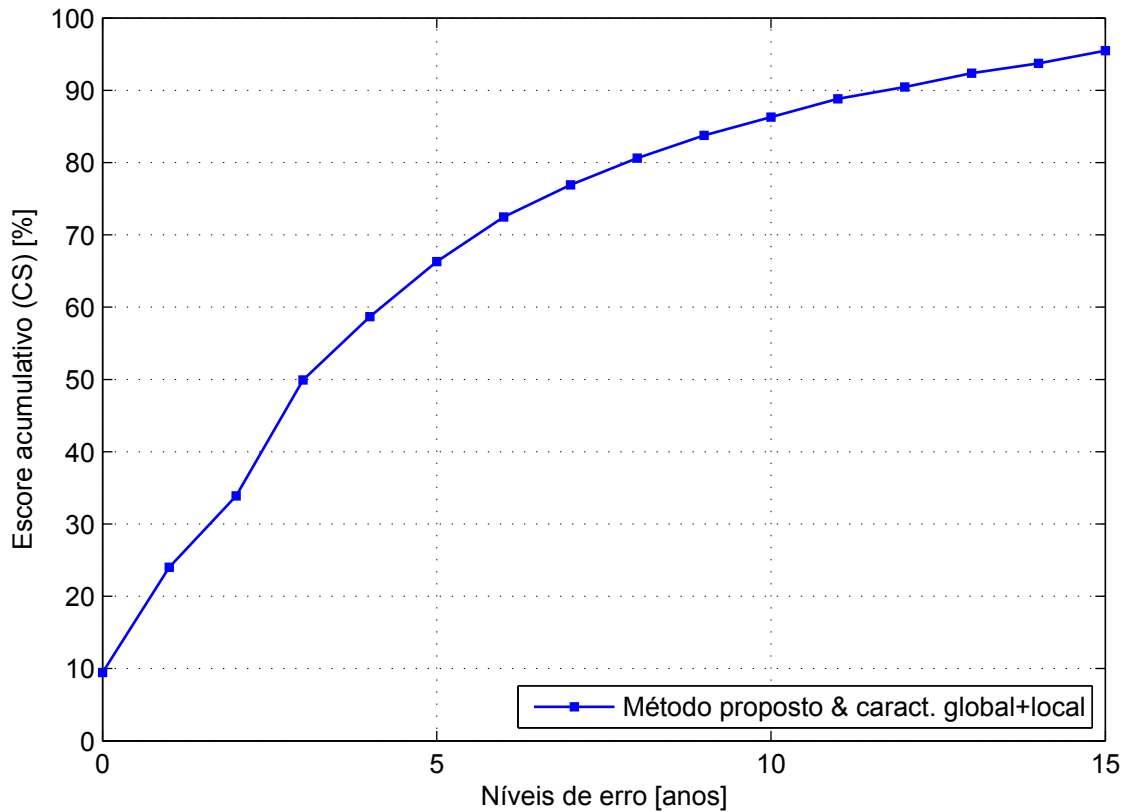


Figura 75: Escore acumulativo ( $CS$ ) da estimativa hierárquica com regiões sobrepostas flexíveis (FG-NET+MORPH).

## 5.7 Estimativa realizada por seres humanos

Para comparar o desempenho do método de estimação etária proposto com a estimativa feita por seres humanos foi realizada uma pesquisa utilizando 250 imagens faciais. Foram selecionadas 25 pessoas de diferentes faixas etárias e gêneros para estimar a idade aparente de 10 imagens faciais cada. As 250 imagens faciais fazem parte da base de dados FG-NET *Aging* e são as mesmas que foram utilizadas para o teste da estimativa hierárquica da idade com regiões sobrepostas flexíveis utilizando o conjunto FG-NET. Para a pesquisa foram utilizados 25 formulários contendo 10 imagens faciais cada. Um exemplo de formulário é mostrado no Apêndice F.

A Tabela 17 mostra a comparação entre o desempenho dos seres humanos e os regressores SVR em estimar a idade aparente em determinadas faixas etárias utilizando o conjunto FG-NET. Os seres humanos tiveram um desempenho superior na faixa etária de 0 a 13 anos, já nas outras faixas etárias os regressores SVR foram melhores em estimar a idade. Na faixa etária de 14 a 21 anos os seres humanos superestimaram diversas idades, obtendo um  $MAE$  de 6,43 anos, enquanto o SVR dessa faixa etária teve mais acertos



obtendo um  $MAE$  de apenas 1,96 anos. Em geral os regressores SVR foram mais precisos em estimar a idade, obtendo um  $MAE$  de 3,39 anos, já os seres humanos tiveram um desempenho inferior com um  $MAE$  de 5,36 anos.

Tabela 17: Comparação entre o desempenho dos seres humanos e os regressores SVR em estimar a idade aparente em determinadas faixas etárias utilizando o conjunto FG-NET

Faixa etária	Estimativa humana	Regressores SVR
	$MAE$	$MAE$
0-13	1,58	2,15
14-21	6,43	1,96
22-39	8,74	5,22
40-69	5,00	4,23
<b>Média geral</b>	5,36	<b>3,39</b>

A Tabela 18 mostra a comparação entre o desempenho dos seres humanos em estimar a idade aparente e o método proposto de estimativa hierárquica da idade com regiões sobrepostas flexíveis utilizando o conjunto FG-NET. As idades aparentes das imagens faciais estimadas por humanos foram divididas em faixas etárias de acordo com a classificação realizada pelo método proposto para ter uma comparação nas mesmas condições. É possível perceber que os seres humanos utilizados na pesquisa tiveram um desempenho inferior ao método computacional proposto, tendo um desempenho levemente superior apenas nas classes 0-28 e 14-22 anos. Apenas uma única imagem foi utilizada na classe 40-69, sendo que o ser humano cometeu um erro de 9 anos e o sistema proposto estimou a idade exata.

Tabela 18: Comparação entre o desempenho dos seres humanos em estimar a idade aparente e o método proposto de estimativa hierárquica da idade com regiões sobrepostas flexíveis utilizando o conjunto FG-NET.

Faixa etária	Estimativa humana	Hierárquico regiões sobrespostas flexíveis
	$MAE$	$MAE$
0-28	3,10	4,08
14-22	6,22	6,33
22-60	7,77	7,52
40-69	9,00	0,09
<b>Média geral</b>	6,52	<b>4,50</b>

A Figura 76 mostra uma comparação dos resultados da estimativa hierárquica da idade com regiões sobrepostas e a estimativa realizada por seres humanos em algumas imagens do conjunto FG-NET. São mostradas as idades reais (*Idade Cronológica*), idades aparentes estimadas por humanos (*Idade Aparente*) e as idades estimadas pelo sistema

(*Idade Estimada*). Em geral, as idades estimadas pelo método proposto são próximas das idades reais. Algumas imagens tiveram erros maiores de estimação, como por exemplo, a imagem facial da mulher de 29 anos a qual foi estimado 17 anos. Porém, percebe-se que existe oclusões causadas por maquiagens, escondendo características importantes para uma estimativa mais precisa. Outro exemplo é a imagem facial da mulher de 47 anos, a qual o sistema subestimou em 13 anos causado talvez pela posição, expressão facial, maquiagem, entre outros fatores. Na imagem facial do homem de 18 anos o ser humano superestimou 32 anos, enquanto o sistema proposto acertou a idade. Isso se deve ao fato da idade de uma pessoa poder ser subestimada ou superestimado devido a sua aparência, como na imagem por exemplo, o bigode faz com que o homem aparente ser mais velho. Já o sistema computacional faz uma análise microtextural da pele e também analisa o formato facial estimando assim uma idade mais precisa.

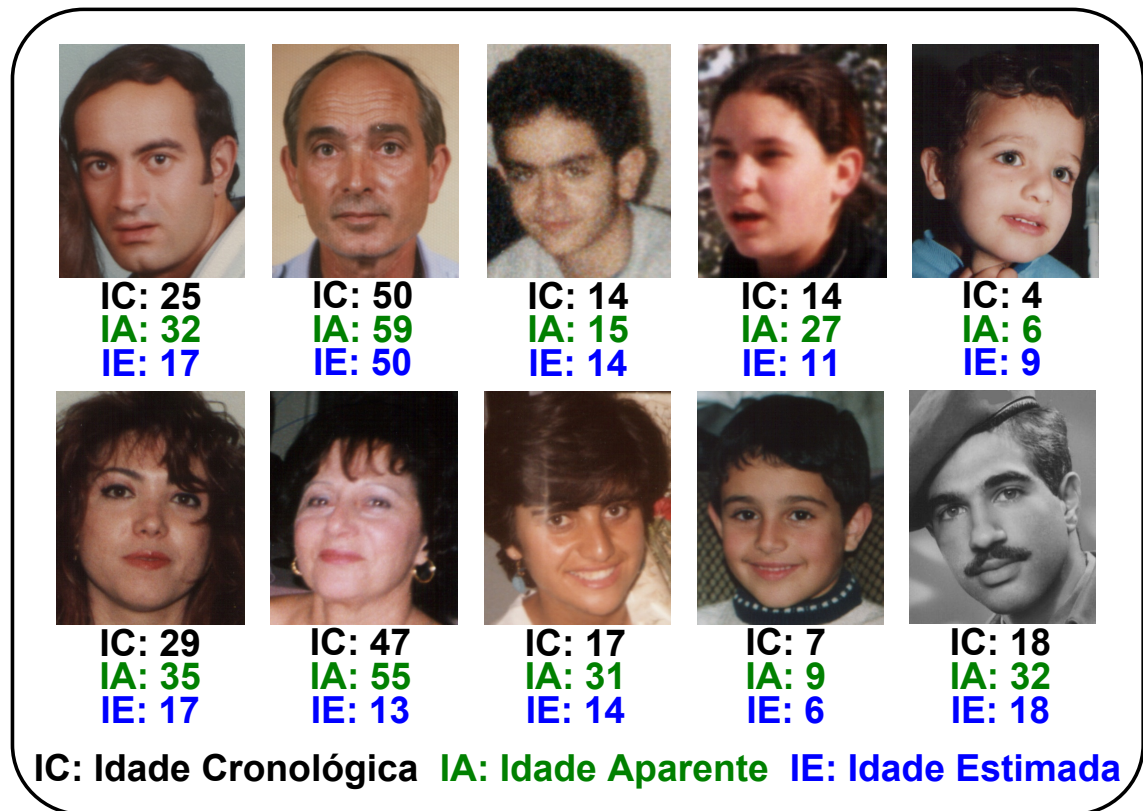


Figura 76: Comparação dos resultados da estimativa hierárquica da idade com regiões sobrepostas e a estimativa realizada por seres humanos em algumas imagens do conjunto FG-NET.

## 5.8 Comparação com outros métodos

### 5.8.1 FG-NET *Aging*

A comparação do método proposto com outros trabalhos reportados no conjunto de dados FG-NET é mostrado na Tabela 19. Kohli et al. [44] e Luu [45] obtiveram um melhor desempenho em seus métodos (3,85 e 4,37 anos de *MAE*, respectivamente), porém, eles utilizaram apenas uma representação holística da face com o método AAM para a extração das características globais o qual não representa muito bem as rugas faciais de faces mais envelhecidas. Para a representação das rugas faciais a extração de padrões de textura em regiões locais devem, como neste trabalho, ser consideradas. Assim como a extração das características em diversas condições de iluminação, expressão, posição e borramento, como também consideradas neste trabalho. O método proposto, estimativa hierárquica com regiões sobrepostas flexíveis utilizando a combinação de características globais (AAM) e locais (LPQ<sub>7×7</sub>) é comparado favoravelmente à literatura independentemente das diferenças nos protocolos experimentais, como por exemplo, o particionamento do conjunto de dados, onde Kohli et al. [44] and Luu [45] particionaram o conjunto de dados em 80% para treinamento e o restante para teste, enquanto neste trabalho foram utilizados 75% para treinamento e 25% para teste. Também, não é possível assegurar que as amostras nos conjuntos de treinamento e teste utilizadas por eles são as mesmas utilizadas neste trabalho. Considerando o protocolo LOPO utilizado, que obteve 4,78 anos de *MAE*, o resultado pode ser comparado com o trabalho de Choi et al. [12], que também utilizaram esse mesmo protocolo, obtendo 4,65 anos de *MAE*, portanto, o método proposto mostra ser competitivo.

Tabela 19: Comparação do método proposto com outros trabalhos reportados no conjunto de dados FG-NET.

Métodos de estimativa etária	<i>MAE</i>
WAS [20]	8,06
AGES [5]	6,77
BM [24]	5,33
RPK [25]	4,95
Luu [45]	4,37
Duong et al. [46]	4,74
Choi et al. [12]	4,65
Kohli et al. [44]	3,85
Estimativa humana	6,52
Método proposto AAM+LPQ <sub>7×7</sub>	<b>4,50</b>
Método proposto AAM+LPQ <sub>7×7</sub> (LOPO)	<b>4,78</b>

### 5.8.2 MORPH *Album 2*

A comparação do método proposto com outros trabalhos reportados no conjunto de dados MORPH *Album 2* é mostrado na Tabela 20. Han et al. [1] obtiveram o melhor desempenho, 5,10 anos de *MAE*, porém utilizaram a versão comercial completa da base de dados MORPH *Album 2* com 78.207 imagens, sendo 10.001 para o conjunto de treinamento e 68.206 para avaliar o desempenho do método. Como representação facial eles utilizaram características biologicamente inspiradas (BIF) extraídas de componentes faciais locais. Apesar do método proposto neste trabalho utilizar apenas 7% (5.474) do total de imagens utilizadas por Han et al., o resultado atingido é muito competitivo.

Tabela 20: Comparação do método proposto com outros trabalhos reportados no conjunto de dados MORPH.

Métodos de estimativa etária	<i>MAE</i>
WAS [20]	9,32
AGES [5]	8,83
Rank [38]	6,49
Han et al. [1]	5,10
Método proposto LPQ <sub>7×7</sub>	<b>5,86</b>

## 6 Conclusões

Neste trabalho foi proposto um novo método para a estimativa hierárquica da idade o qual é composto por uma etapa de classificação em faixas etárias utilizando SVM e uma segunda etapa para a estimativa específica da idade utilizando regressores SVR de acordo com a classe selecionada na primeira etapa. Para compensar os erros gerados na etapa de classificação causados pelas margens fixas entre as classes os regressores SVR foram treinados com regiões sobrepostas flexíveis entre as faixas etárias. Para a representação das características faciais relacionadas ao envelhecimento foram utilizadas características globais e locais. As características globais foram extraídas através de um modelo estatístico da aparência e formato facial gerado pela técnica AAM, enquanto as características locais foram extraídas utilizando diversas combinações e técnicas como Gabor *wavelets*, LBP e LPQ após um pré-processamento das imagens e localização das regiões de interesse da face. Após a combinação das características globais e locais através do método *feature level fusion* para uma representação mais completa da face e suas características relacionadas à idade, foi realizada uma redução da dimensionalidade das características utilizando PCA tendo um ganho em desempenho e eficiência computacional.

Os conjuntos de dados FG-NET *Ageing* e MORPH *Album 2*, assim como uma combinação de ambos foram utilizados para avaliar o desempenho do sistema proposto e também de outros métodos para efeitos de comparação. O método proposto, estimativa hierárquica da idade com regiões sobrepostas flexíveis, avaliado no conjunto de dados FG-NET teve um desempenho médio de 6,53 anos de *MAE*, superior a todos os outros métodos utilizados nessa pesquisa como a estimativa *single-level*, estimativa hierárquica com margens fixas e estimativa hierárquica com regiões sobrepostas fixas. O melhor desempenho em termos de *MAE*, 4,50 anos, foi atingido utilizando a combinação de características globais e locais AAM+LPQ<sub>7×7</sub>. Utilizando essa mesma combinação de características, o desempenho obtido através do protocolo LOPO foi de 4,78 anos, onde garante que uma pessoa não esteja no conjunto de teste e treinamento ao mesmo tempo eliminando assim a dependência dos dados nos resultados do experimento. Utilizando o

conjunto MORPH, o melhor desempenho em termos de  $MAE$  foi de 5,86 anos utilizando as características locais  $LPQ_{7 \times 7}$ . Com a combinação dos conjuntos de dados FG-NET e MORPH o melhor desempenho obtido foi de 5,20 anos de  $MAE$ , utilizando a combinação de características globais e locais AAM+ $LPQ_{7 \times 7}$ . Sendo possível concluir que as características locais extraídas utilizando o descritor LPQ foram as mais discriminantes em ambos os conjuntos de dados, MORPH e FG-NET assim como na combinação FG-NET+MORPH mostrando os melhores resultados em termos de  $MAE$ . O descritor de textura LPQ mostrou-se muito robusto não apenas ao borramento mas também a outros desafios como variações na iluminação e expressão facial muito presente nos conjuntos de dados utilizados. Apesar da simplicidade e robustez o LPQ se mostrou muito discriminativo produzindo resultados muito satisfatórios para a estimativa da idade superando os outros métodos utilizados em outros trabalhos como Gabor *wavelets* e LBP.

Também foi comparado neste trabalho o desempenho do método de estimação etária proposto com a estimativa da idade aparente feita por seres humanos. Para isso foi realizada uma pesquisa utilizando 250 imagens faciais do conjunto FG-NET e selecionadas 25 pessoas de diferentes faixas etárias e gêneros para estimar a idade aparente de 10 imagens faciais cada. Com a pesquisa foi possível concluir que os seres humanos tiveram um desempenho inferior ao método proposto, mostrando que sistemas computacionais são mais precisos nas estimativa etária. Comparações com outros trabalhos reportados no conjunto de dados FG-NET e MORPH também foram realizadas, onde Kohli et al. [44] e Luu [45] obtiveram um melhor desempenho no conjunto FG-NET com seus métodos (3,85 e 4,37 anos de  $MAE$ , respectivamente), porém, eles utilizaram apenas uma representação holística da face com o método AAM para a extração das características globais o qual não representa muito bem as rugas faciais de faces mais envelhecidas. Para a representação das rugas faciais a extração de padrões de textura em regiões locais devem, como realizadas nesse trabalho, ser consideradas. Assim como a extração das características em diversas condições de iluminação, expressão, posição e borramento, como também consideradas nesse trabalho. Já no conjunto MORPH, Han et al. [1] obtiveram o melhor desempenho, 5,10 anos de  $MAE$ , porém utilizaram a versão comercial completa da base de dados MORPH *Album 2* com 78.207 imagens, sendo 10.001 para o conjunto de treinamento e 68.206 para avaliar o desempenho do método. Portanto, o método proposto, seja avaliado no conjunto FG-NET ou MORPH é comparado favoravelmente à literatura independentemente das diferenças nos protocolos experimentais.

Alguns fatores que afetaram o desempenho do método proposto para estimativa da idade são apresentados a seguir:

1. Rugas faciais são geralmente removidas pelo fotógrafo ou devido ao uso de cosméticos principalmente nas mulheres que utilizam maquiagem e cremes anti-rugas, as quais reduzem o desempenho dos extratores de características;
2. Algumas imagens dos conjuntos FG-NET e MORPH foram tiradas utilizando fontes luminosas muito brilhantes fazendo com que importantes características faciais relacionadas ao envelhecimento fossem perdidas;
3. Algumas oclusões como o uso de óculos, bigodes e barbas em algumas imagens também fazem com que algumas características importantes sejam perdidas.
4. Os 68 pontos de referência (*landmarks*) do conjunto MORPH foram extraídos automaticamente, portanto existem alguns erros de posicionamento de componentes faciais que prejudicam o desempenho da extração das características locais.
5. Algumas regiões de interesse foram recortadas com elementos faciais não desejados, como por exemplo, parte da sobrancelha, olho e cabelos.

## 6.1 Contribuições

As principais contribuições deste trabalho podem ser descritas da seguinte forma:

- Apresentação de um novo método de estimativa hierárquica da idade com regiões sobrepostas flexíveis.
- Pesquisa utilizando diversas técnicas para a extração de características, assim como várias combinações e mudanças de parâmetros.
- Utilização da técnica LPQ para a extração de características locais relacionadas ao envelhecimento, sendo robusto não apenas ao borramento mas também a outros fatores como variações na iluminação e expressão facial muito presente nos conjuntos de dados utilizados.
- Utilização do conjunto de dados MORPH *Album 2* sendo pouco explorado nos trabalhos anteriores para avaliar o desempenho do método proposto.
- Combinação das bases FG-NET e MORPH afim de verificar a robustez do método proposto.

## 6.2 Trabalhos futuros

As seguintes considerações podem ser tomadas em trabalhos futuros:

- Utilização de outras bases de dados para a avaliação do método como por exemplo, a base FACES [47], a qual contém imagens faciais controladas de 171 pessoas jovens, adultas e idosas com 2 conjuntos de 6 diferentes expressões faciais para cada pessoa totalizando em 2.052 imagens de homens e mulheres de 19 a 80 anos.
- Esforços na criação de uma base de dados pública contendo diversas idades de um mesmo indivíduo, sendo o ideal uma base que contém diversos indivíduos de ambos os gêneros, diversas etnias e contendo pelo menos uma imagem facial para toda faixa etária, por exemplo, de 0 a 80 anos.
- Investigar outras técnicas para a extração de características, como por exemplo, o robusto descritor de textura WLD (*Weber's Law Descriptor*) proposto por Chen et al. [48].
- Utilização de mais *landmarks* para a extração de características globais utilizando AAM para cobrir a região da testa, não coberta pelos 68 pontos de referência utilizados neste trabalho.
- Investigar a técnica de aprendizagem profunda (*Deep Learning*) aplicada a estimativa da idade.
- Investigar novas métricas para a avaliação do desempenho.
- Desenvolvimento de um protótipo funcional.



## Referências

- [1] H. Han, C. Otto, and A. K. Jain, “Age estimation from face images: Human vs. machine performance,” *The 6th IAPR International Conference on Biometrics (ICB)*, pp. 1–8, 2013.
- [2] Y. Fu, G. Guo, and T. S. Huang, “Age synthesis and estimation via faces: A survey,” *IEEE T Pattern Anal*, vol. 32, pp. 1955–1976, 2010.
- [3] N. Ramanathan and R. Chellappa, “Modeling age progression in young faces,” in *IEEE CVPR*, 2006, pp. 387–394.
- [4] Y. H. Kwon and N. D. V. Lobo, “Age classification from facial images,” *Comp Vis Image Und*, vol. 74, pp. 1–21, 1999.
- [5] X. Geng, Z. H. Zhou, and K. Smith-Miles, “Automatic age estimation based on facial aging patterns,” *IEEE T Pattern Anal*, vol. 29, pp. 2234–2240, 2007.
- [6] Y. Fu, Y. Xu, and T. S. Huang, “Estimating human ages by manifold analysis of face pictures and regression on aging features,” in *IEEE Conf. Multimedia and Expo*, 2007, pp. 1383–1386.
- [7] Y. Fu and T. S. Huang, “Human age estimation with regression on discriminative aging manifold,” in *IEEE Trans. Multimedia*, 2008, pp. 578–584.
- [8] A. Stone, “The Aging Process of the Face & Techniques of Rejuvenation,” [http://www.aaronstonemd.com/Facial\\_Aging\\_Rejuvenation.shtm](http://www.aaronstonemd.com/Facial_Aging_Rejuvenation.shtm), 2010.
- [9] T. F. Cootes, D. Cooper, C. J. Taylor, and J. Graham, “Active shape models: Their training and application,” *Computer Vision and Image Understanding*, vol. 61, pp. 38–59, 1995.
- [10] M. Nixon, M. S. Nixon, and A. S. Aguado, “Feature extraction and image processing for computer vision,” p. 609, 2012.
- [11] T. Ahonen and E. Rahtu, “Recognition of blurred faces using local phase quantization,” *IEEE CVPR*, pp. 1–4, 2008.
- [12] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, “Age estimation using a hierarchical classifier based on global and local facial features,” *Patt Recog*, vol. 44, pp. 1262–1281, 2011.
- [13] M. Pietikainen and T. Maenpaa, “Texture analysis with local binary patterns,” pp. 1–20, 2004.
- [14] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar, “Describable visual attributes for face verification and image search,” *IEEE Transactions on PAMI*, pp. 1962–1977, 2011.

- [15] S. Thakur and L. Verma, "Identification of face age range group using neural network," *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, pp. 250–254, 2012.
- [16] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 34, pp. 621–628, 2004.
- [17] M. Rhodes, "Age estimation of faces: A review," *Applied Cognitive Psychology*, vol. 12, pp. 1–12, 2009.
- [18] N. Hewahi, A. Olwan, and N. Tubeel, "Estimation based on neural networks using face features," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 1, pp. 61–67, 2010.
- [19] L. Farkas, "Anthropometry of the head and face," *Raven Press*, 1994.
- [20] A. Lanitis, C. Taylor, and T. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE T Pattern Anal*, vol. 24, pp. 442–455, 2002.
- [21] G. Guo, Y. Fu, T. S. Huang, and C. Dyer, "Locally adjusted robust regression for human age estimation," in *Proc. IEEE Workshop Applications of Computer Vision*, 2008.
- [22] J. Hayashi, M. Yasumoto, H. Ito, Y. Niwa, and H. Koshimizu, "Age and gender estimation from facial image processing," *SICE Ann. Conf.*, vol. 1, pp. 13–18, 2002.
- [23] H. Fukai, H. Takimoto, Y. Mitsukura, and M. Fukumi, "Apparent age estimation system based on age perception," *Proc. SICE Ann. Conf.*, pp. 2808–2812, 2007.
- [24] S. Yan, H. Wang, T. Huang, Q. Yang, and X. T. Xiaoou, "Ranking with uncertain labels," in *Int'l Cong. Math. Educ.*, 2007, pp. 96–99.
- [25] S. Yan, X. Zhou, M. Hasegawa-Johnson, and T. S. Huang, "Regression from patch-kernel," in *IEEE CVPR*, 2008, pp. 1–8.
- [26] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio inspired features," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2009, pp. 112–119.
- [27] T. Kanno, M. Akiba, Y. Teramachi, H. Nagahashi, and T. Agui, "Classification of age group based on facial images of young males by using neural networks," in *IEICE Trans. Information and Systems*, 2001, pp. 1094–1101.
- [28] J. Suo, T. Wu, S. Zhu, S. Shan, X. Chen, and W. Gao, "Sparse features for age estimation using hierarchical face model," in *Proc. IEEE Conf. Automatic Face and Gesture Recognition*, 2008.
- [29] B. Ni, Z. Song, and S. Yan, "Web image mining towards universal age estimator," in *Proc. ACM Multimedia*, 2009.
- [30] C. Lin, D. Li, J. Lai, M. Han, and J. Chang, "Automatic age estimation system for face images," *International Journal of Advanced Robotic Systems*, pp. 1–9, 2012.

- [31] I. Daubechies, “The wavelet transform, time-frequency localization and signal analysis,” *IEEE Transactions on Information Theory*, vol. 36, pp. 961–1005, 1990.
- [32] T. Ojala, M. Pietikainen, and T. Maenpaa, “A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification,” *Pattern Recognition ICAPR*, pp. 1–10, 2001.
- [33] T Ojala, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Transactions on PAMI*, vol. 24, pp. 971–987, 2002.
- [34] V. Ojansivu and J. Heikkil, “Using local phase quantization,” in *Int’l Conf. Image Signal Process*, 2008, pp. 236–243.
- [35] S. Haykin, *Neural Networks - A Comprehensive Foundation - Second Edition*, Prentice Hall, 2005.
- [36] “The fg-net aging database,” <http://sting.cycollege.ac.cy/alanitis/fgnetaging/index.htm>, 2010.
- [37] K. Ricanek and T. Tesafaye, “Morph: A longitudinal image database of normal adult age-progression,” *IEEE International Conference on Automatic Face and Gesture*, pp. 341–345, 2006.
- [38] K. Chang, C. Chen, and Y. Hung, “A ranking approach for human ages estimation based on face images,” *20th International Conference on Pattern Recognition*, pp. 3396–3399, 2010.
- [39] A. Ross and R. Govindarajan, “Feature level fusion in biometric systems,” in *Proc. of the Biometric Cons. Conf.*, 2004, pp. 1–2.
- [40] C. Chang and C. Lin, “Libsvm: A library for support vector machines,” pp. 1–39, 2013.
- [41] R. Fan, X. Wang, and C. Lin, “Liblinear: A library for large linear classification,” *Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2012.
- [42] T. Cootes, G. Edwards, and C. Taylor, “Active appearance models,” *IEEE T Pattern Anal*, vol. 23, pp. 681–685, 2001.
- [43] A. Vedrana and A. Renee, “Active shape and appearance models,” *IT University of Copenhagen*, p. 76, 2005.
- [44] S. Kohli, S. Prakash, and P. Gupta, “Hierarchical age estimation with dissimilarity-based classification,” *Neurocomputing*, vol. 120, pp. 164–176, 2013.
- [45] K. Luu, K. Ricanek, T. D. Bui, and C. Y. Suen, “Age estimation using active appearance models and support vector machine regression,” in *IEEE Int’l Conf. Biometrics*, 2009, pp. 1–5.
- [46] C. N. Duong, K. G. Quach, K. Luu, H. B. Le, and K. Ricanek, “Fine tuning age estimation with global and local facial features,” in *IEEE ICASSP*, 2011, pp. 2032–2035.

- [47] N. C. Ebner, M. Riediger, and U. Lindenberger, “Faces - a database of facial expressions in young, middle-aged, and older women and men: development and validation.,” *Behavior research methods*, vol. 42, pp. 351–362, 2010.
- [48] C. Chen, Y. Chang, K. Ricanek, and Y. Wang, “Face age estimation using model selection,” in *IEEE Int’l Conf. Biometrics*, 2010, pp. 93–99.

## APÊNDICE A – Análise *local binary patterns*

A Figura 77 mostra o histograma dos códigos  $LBP_{8,2}^{u2}$  para a região do canto direito da boca. Foi utilizado um exemplo para cada faixa etária: bebê de 2 anos, jovem de 17 anos, adulto de 28 anos e idoso de 69 anos. Conforme uma pessoa vai envelhecendo linhas nos cantos da boca, linhas de marionete, sulcos e manchas começam a aparecer. Analisando o histograma do bebê de 2 anos percebe-se que não possui pontos brilhantes (*spots*), onde todos os pontos circundantes possuem um valor mais baixo que o pixel central. O código LBP para essa representação é o 1. Devido a pele ser de um bebê, normalmente as manchas devem ser mínimas como mostrado no histograma, pois normalmente tais manchas são relacionadas ao envelhecimento. Outro código LBP importante para essa faixa etária é o 58 o qual representa um ponto escuro com os pontos circundantes mais brilhantes, ou seja, uma região com textura suave. Em teoria, deveria ser o código mais predominante do histograma uma vez que a pele de um bebê tem uma textura suave, porém, apesar do operador LBP reduzir com sucesso ruídos utilizando padrões uniformes, ainda é sensível a ruídos na imagem. Um pequeno ruído pode causar diferenças nos pixels fazendo com que o código LBP seja codificado com outro código.

Idealmente uma região lisa deveria ser codificada como "11111111", mas com algum ruído poder ser codificado como um padrão não-uniforme perdendo tal característica de suavidade. Nos histogramas dos indivíduos de 17 e 28 anos algumas manchas já podem ser percebidas pela análise dos códigos LBP, representadas pelo código 1. Também começam a surgir mais códigos relacionados a términos de linhas (códigos 9 e 24), pontiagudos (código 10) e outros códigos que representam linhas faciais (código 8, por exemplo). O canto direito da boca do indivíduo de 69 anos possui um histograma com predominância dos códigos 14 e 19, por exemplo, os quais representam as rugas da região do canto da boca assim como manchas e outras linhas em diversas direções representadas pelos códigos 27 e 43.

A Figura 78 mostra o histograma dos códigos  $LBP_{8,2}^{u2}$  para a região das linhas da bochecha direita dos mesmos indivíduos descritos anteriormente. A pele da região da

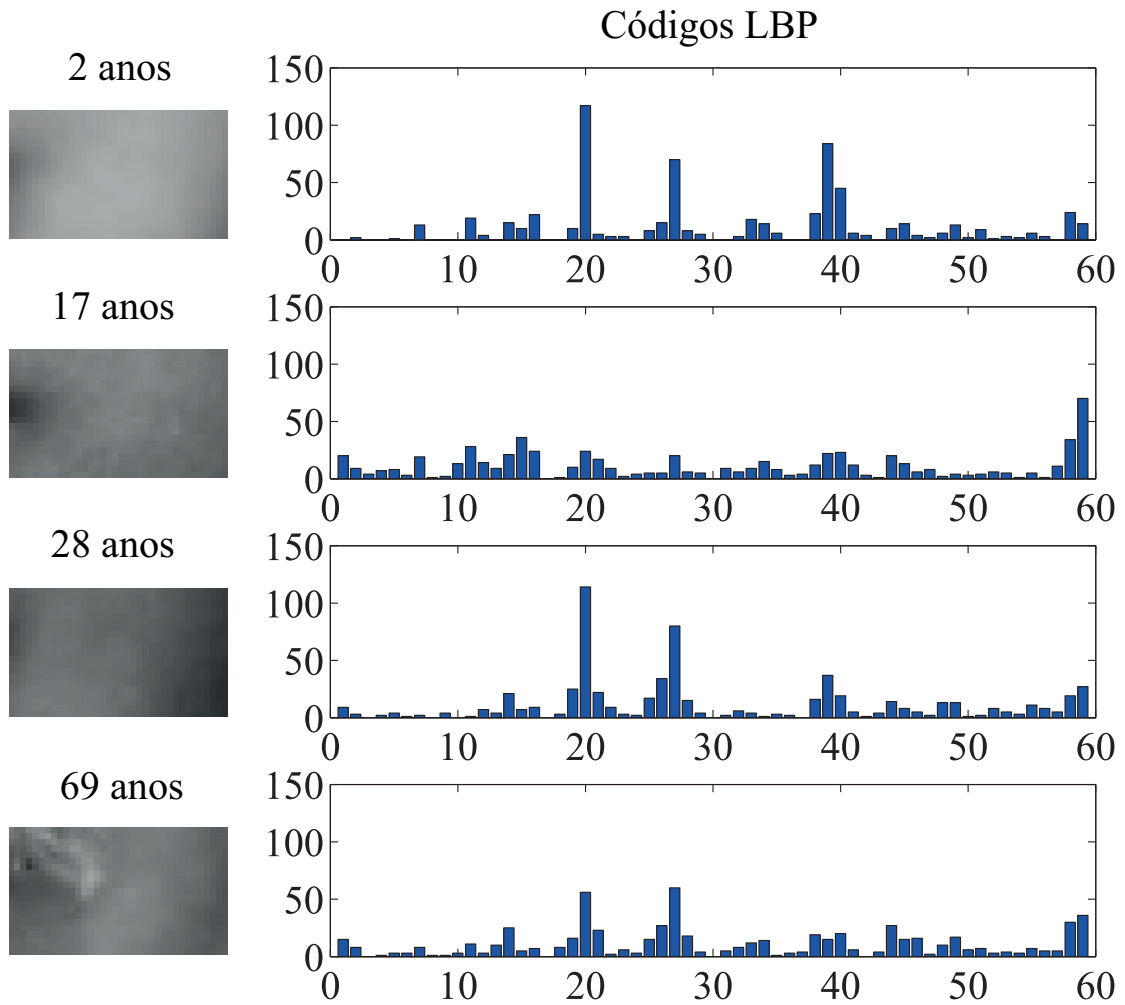


Figura 77: Histogramas dos códigos  $LBP_{8,2}^u$  para a região do canto direito da boca para 2, 17, 28 e 69 anos respectivamente (FG-NET).

bochecha de um bebê provavelmente será uma pele lisa e suave como mostra o código 57 no histograma. Também, com poucas manchas claras, como o previsto. Os códigos 48 e 49, passando de 200, são linhas ruidosas com um certo grau de curvatura causadas pela região abaixo do olho. O histograma do indivíduo de 17 anos possui um pouco mais de códigos relacionados a manchas. Assim como o bebê, a pele de uma pessoa de 17 anos normalmente possui uma textura suave nas bochechas. Há também um número maior de padrões não-uniformes, código 59. No indivíduo de 28 anos o histograma possui mais códigos relacionados a bordas e cantos, como por exemplo, o código 14 e 19. No histograma da bochecha do indivíduo de 69 anos é possível perceber um número maior de códigos relacionados a manchas causadas pelo envelhecimento. Padrões de rugas pontiagudas também são mais proeminentes definidos pelo código 7. Também, um número maior dos códigos 10 e 24 que representam padrões de cantos e bordas, respectivamente.

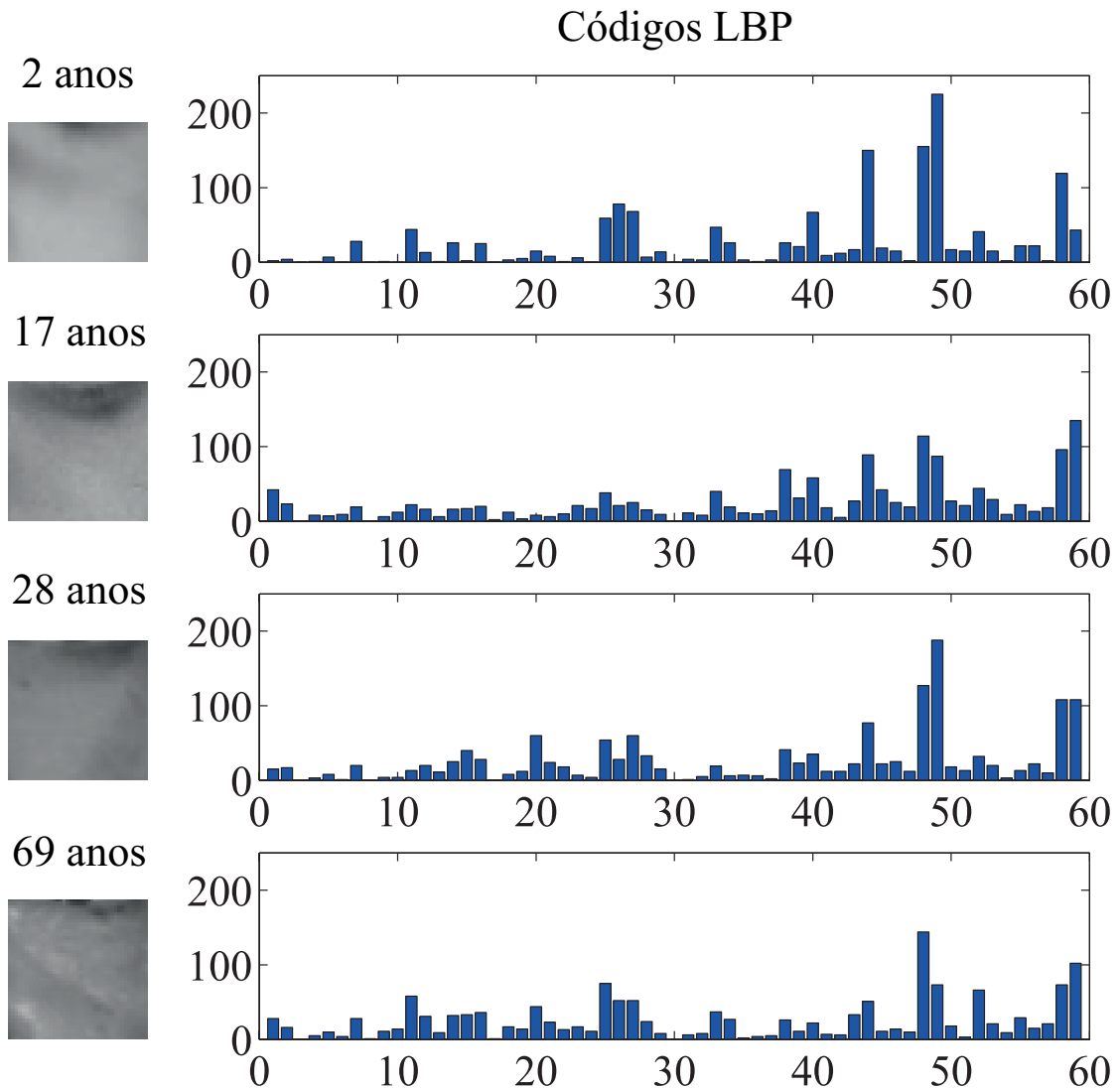


Figura 78: Histogramas dos códigos  $LBP_{8,2}^{u2}$  para a região das linhas da bochecha direita para 2, 17, 28 e 69 anos respectivamente (FG-NET).

A Figura 79 mostra o histograma dos códigos  $LBP_{8,2}^{u2}$  para a região das linhas do queixo dos mesmos indivíduos descritos anteriormente. A pele da região do queixo de um bebê provavelmente será uma pele lisa e suave. Também, com poucas manchas claras, como o previsto pelo histograma. Os códigos 11 e 15, passando de 400, são devido a borda que divide o queixo do pescoço causado pelo fato do bebê possuir um formato menor da cabeça, assim a região de interesse é cortada capturando uma região do pescoço. O histograma do indivíduo de 17 anos possui um pouco mais de códigos relacionados a manchas. Assim como o bebê, a pele de uma pessoa de 17 anos normalmente possui uma textura suave no queixo, porém, já são proeminentes algumas linhas faciais. Há também um número maior de padrões não-uniformes, código 59. No indivíduo de 28 anos o histograma possui mais códigos relacionados a bordas, como por exemplo, o código 24 e

26. No histograma do queixo do indivíduo de 69 anos é possível perceber um número maior de códigos relacionados a manchas causadas pelo envelhecimento. Padrões de códigos relacionados a bordas e cantos também são mais distribuídos, definidos pelo código 7, 11 e 14. Também, um número maior do código 59 que representam padrões não-uniformes os quais definem várias curvaturas de linhas faciais.

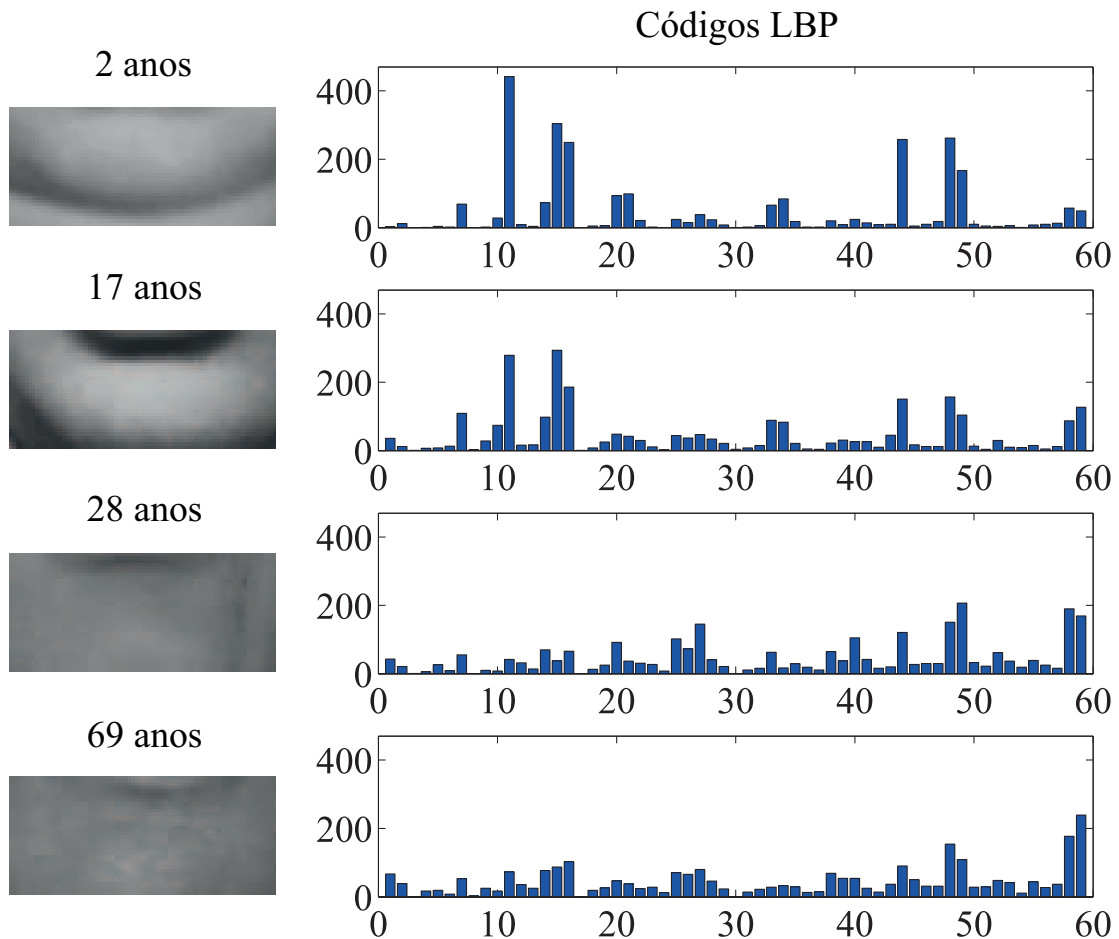


Figura 79: Histogramas dos códigos  $LBP_{8,2}^{u2}$  para a região das linhas do queixo para 2, 17, 28 e 69 anos respectivamente (FG-NET).

A Figura 80 mostra o histograma dos códigos  $LBP_{8,2}^{u2}$  para a região do topo do nariz dos mesmos indivíduos descritos anteriormente. A pele da região do topo do nariz de um bebê provavelmente será uma pele lisa e suave, como mostra o código 58 no histograma. Também, com poucas manchas claras. Os códigos 20 e 26 são devido as bordas do formato do nariz causado pelo fato do bebê possuir um nariz menor. O histograma do indivíduo de 17 anos possui um pouco mais de códigos relacionados a manchas, pois o nariz é uma região muito exposta a climas extremos (sol, chuva, frio etc.) sendo que prejudica a pele causando tais manchas. Há também um número maior de padrões não-uniformes, código 59. No indivíduo de 28 anos o histograma é similar



ao do indivíduo de 17 anos. No histograma do indivíduo de 69 anos é possível perceber um número maior de códigos relacionados a manchas causadas pelo envelhecimento no topo do nariz. Uma predominância de bordas, ou rugas, definidas pelos códigos 20, 26 e outros são percebidas. E também, um número maior do código 58 que representam padrões não-uniformes os quais definem várias curvaturas de linhas faciais.

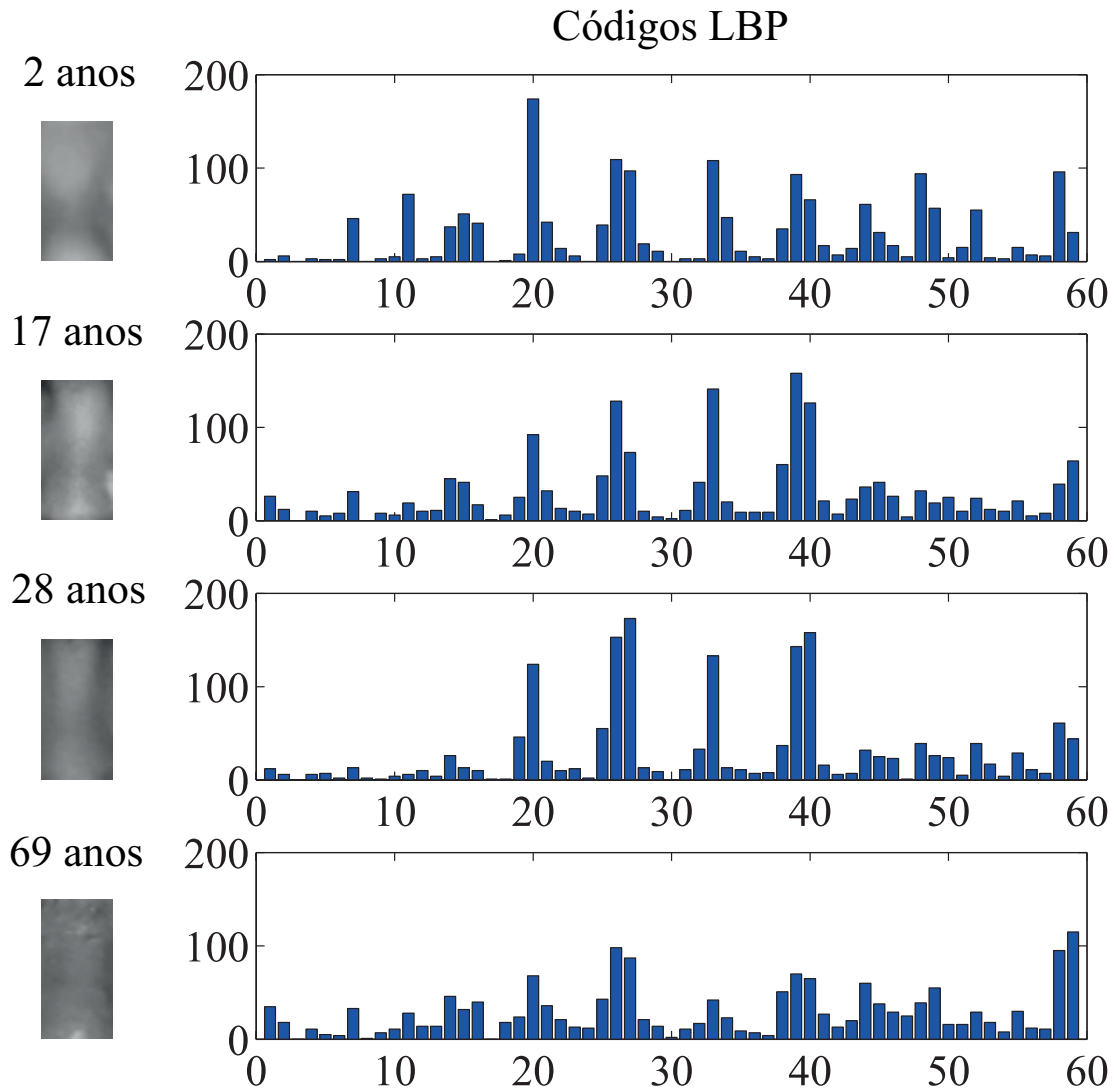


Figura 80: Histogramas dos códigos  $LBP_{8,2}^{u2}$  para a região do topo do nariz para 2, 17, 28 e 69 anos respectivamente (FG-NET).

A Figura 81 mostra o histograma dos códigos  $LBP_{8,2}^{u2}$  para a região das linhas nasolabial direita. Foi utilizado um exemplo para cada faixa etária: bebê de 2 anos, jovem de 17 anos, adulto de 28 anos e idoso de 69 anos. A região nasolabial é uma região onde as rugas aparecem proeminentes formando o bigode chinês, uma região importante para a extração de características relacionadas ao envelhecimento. A pele da região nasolabial de um bebê de 2 anos provavelmente não terá o bigode chinês, será uma pele lisa e

suave como mostra o código 58 no histograma sendo que chega próximo de 150 códigos encontrados. Também, com poucas manchas claras, como o previsto. Os histogramas dos indivíduos de 17 anos e 28 anos possuem um pouco mais de códigos relacionados a manchas. Assim como o bebê, a pele de uma pessoa de 17 anos e 28 anos normalmente é suave com linhas nasolabial pouco proeminentes em expressões neutras. Porém, já é possível verificar que existem tais linhas e cantos justamente pela predominância do código 13 o qual é relacionado a borda de uma linha nasolabial. No histograma do indivíduo de 69 anos é possível perceber um número maior de códigos relacionados a manchas. Códigos relacionados a bordas nasolabial e cantos também são proeminentes como os códigos 7, 32 e 33.

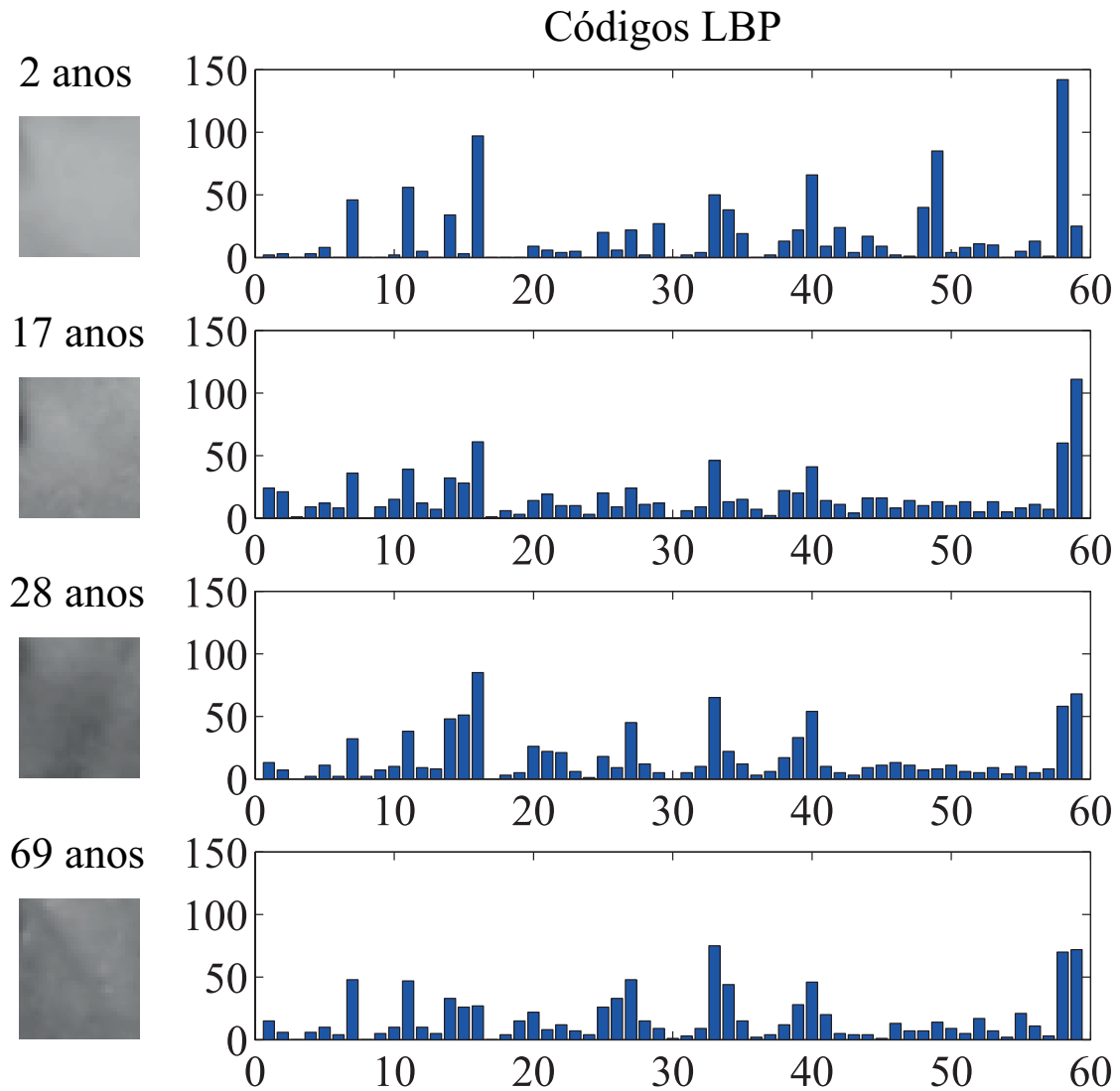


Figura 81: Histogramas dos códigos  $LBP_{8,2}^{u2}$  para a região das linhas nasolabial direitas para 2, 17, 28 e 69 anos respectivamente (FG-NET).

## APÊNDICE B – *Análise local phase quantization*

É possível perceber em todos os histogramas mostrados nas Figuras 82 a 86 que os valores dos códigos LPQ são geralmente maiores para indivíduos jovens (2, 17 e 28 anos), pois possuem graus de orientação de fases similares, normalmente em um mesmo sentido, devido a possuírem uma textura de pele mais suave. No caso do indivíduo de 68 anos a textura da pele possui rugas em diversas orientações como é possível perceber pelo gráfico de orientações das figuras, onde os vetores unitários que demonstram a orientação das características estão em diversos sentidos. Também, conforme a pele vai ficando com uma textura mais envelhecida o histograma de códigos LPQ vai tendo uma distribuição mais uniforme com valores baixos devido a possuir mais graus de orientação das fases causado pelas rugas.

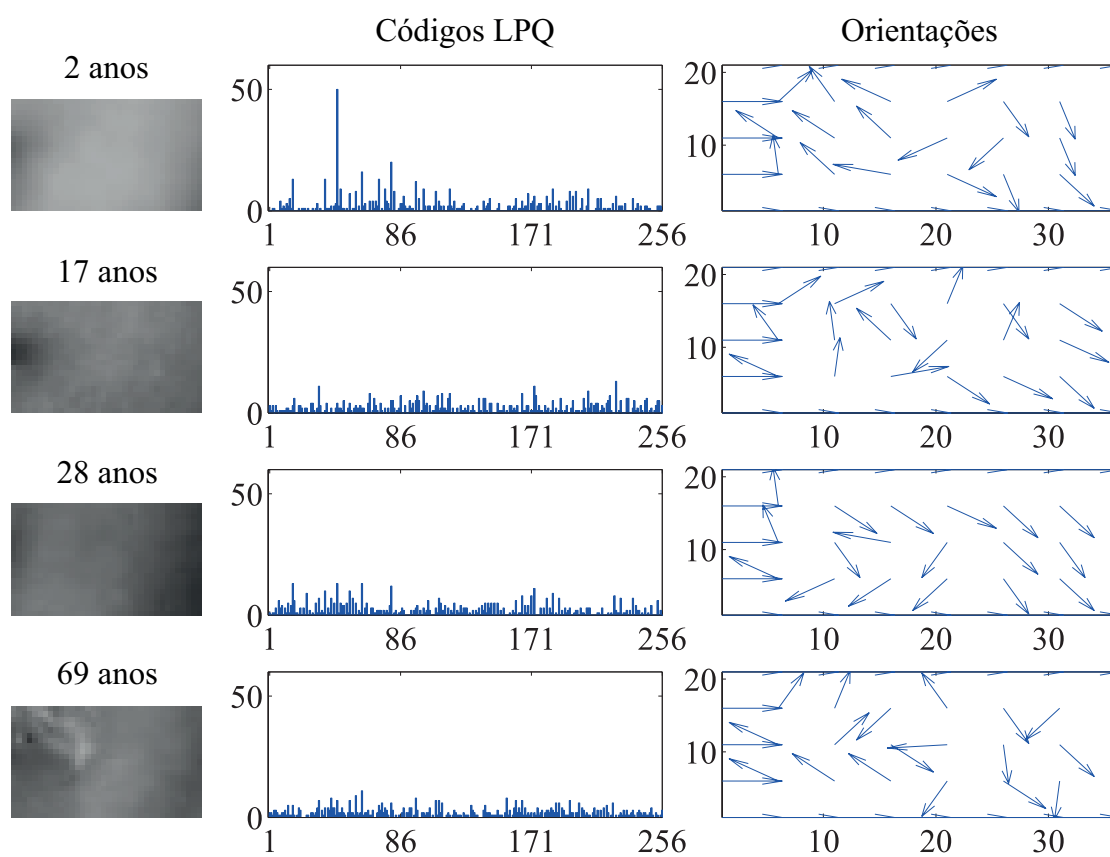


Figura 82: Histogramas dos códigos  $LPQ_{5 \times 5}$  para a região do canto direito da boca e suas respectivas orientações para 2, 17, 28 e 69 anos respectivamente (FG-NET).

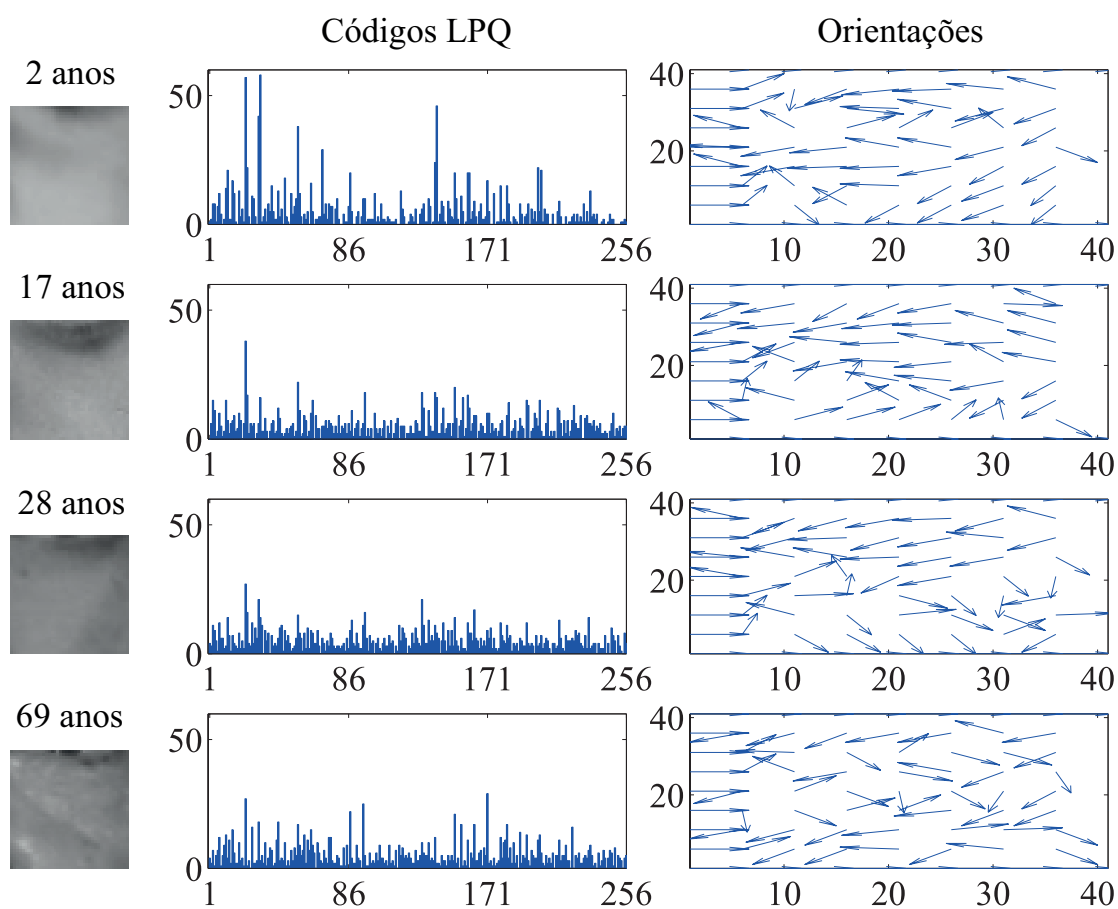


Figura 83: Histogramas dos códigos  $LPQ_{5 \times 5}$  para a região das linhas da bochecha direita e suas respectivas orientações para 2, 17, 28 e 69 anos respectivamente (FG-NET).

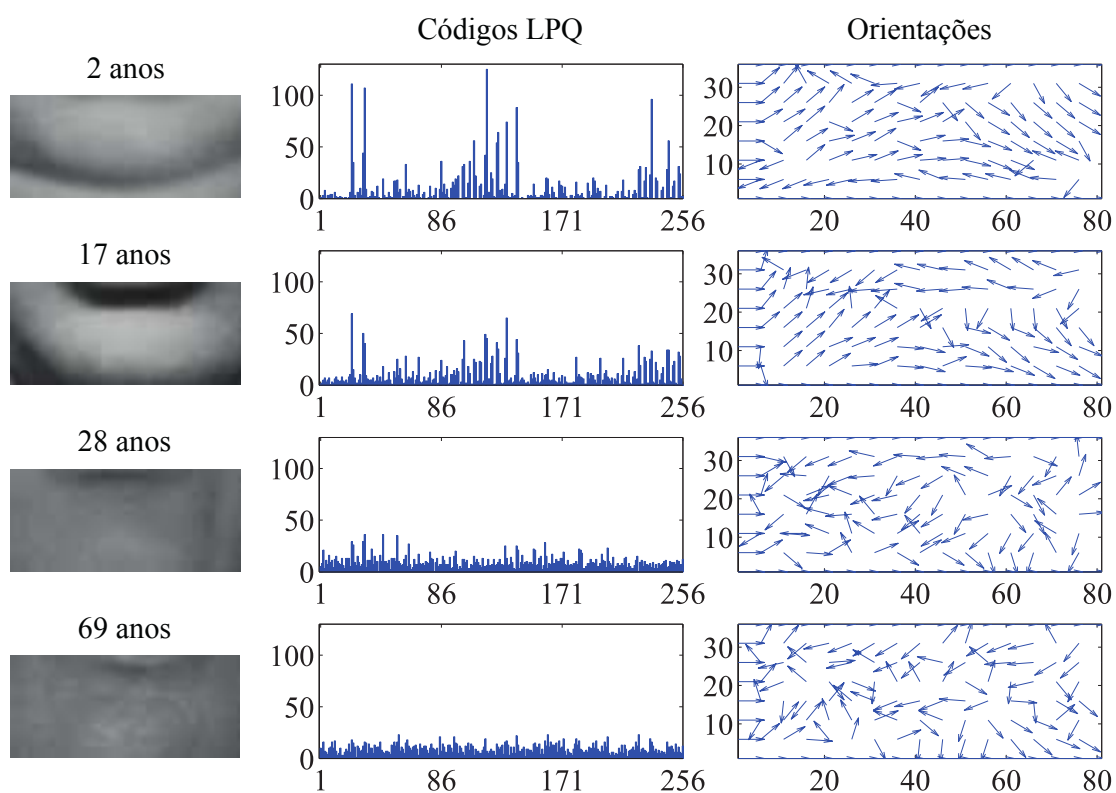


Figura 84: Histogramas dos códigos  $LPQ_{5 \times 5}$  para a região das linhas do queixo e suas respectivas orientações para 2, 17, 28 e 69 anos respectivamente (FG-NET).

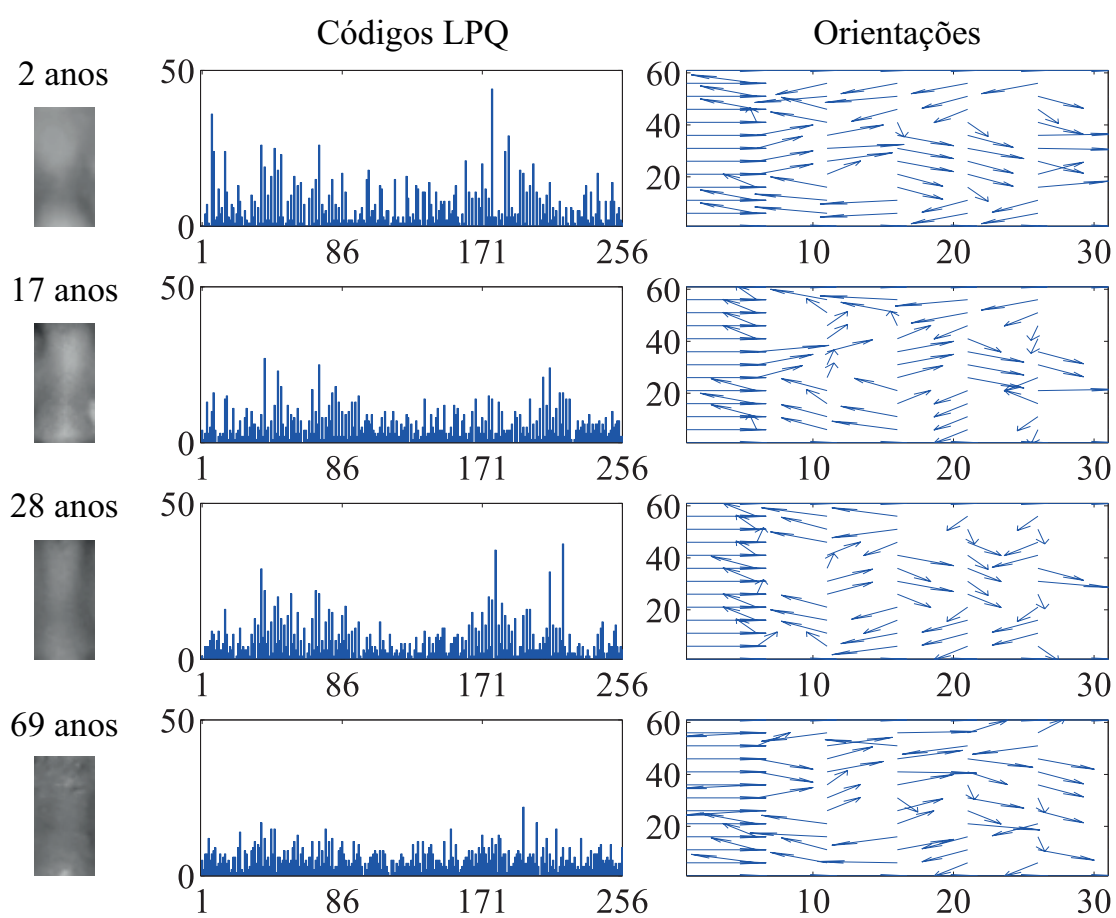


Figura 85: Histogramas dos códigos  $LPQ_{5 \times 5}$  para a região do topo do nariz e suas respectivas orientações para 2, 17, 28 e 69 anos respectivamente (FG-NET).

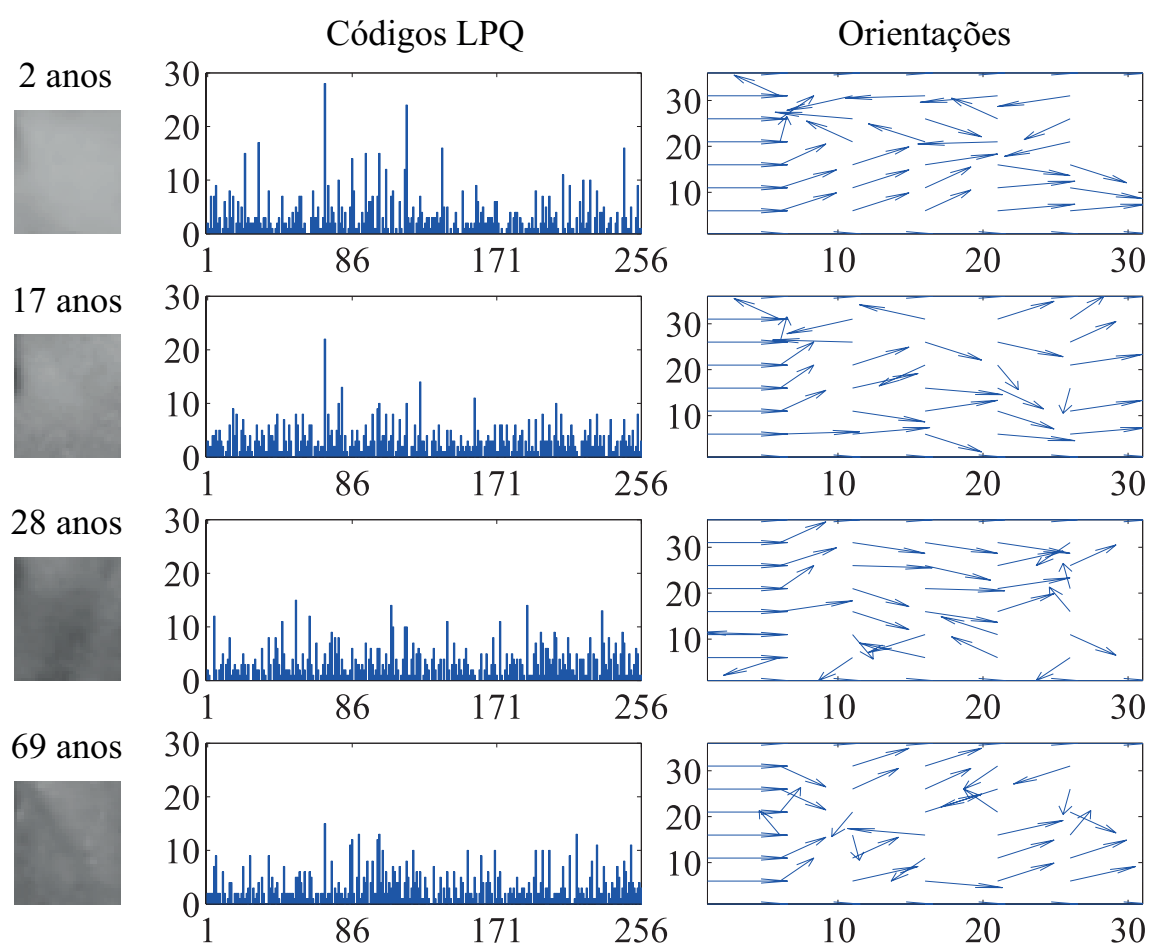


Figura 86: Histogramas dos códigos  $LPQ_{5 \times 5}$  para a região das linhas nasolabial direita e suas respectivas orientações para 2, 17, 28 e 69 anos respectivamente (FG-NET).



## APÊNDICE C – Estimativa hierárquica por décadas

A estimativa hierárquica por décadas também foi avaliada neste trabalho, onde são definidas faixas etárias mais restritas resultando em um número maior de classes para a etapa da classificação em faixas etárias e também de regressores para a estimativa específica da idade. Devido o conjunto de dados FG-NET possuir uma faixa etária de 0 a 69 anos foram definidas 7 classes com faixas etárias espaçadas em aproximadamente 10 anos cada (décadas) para a etapa de classificação e regressão. As faixas etárias e as distribuições dos dados de treinamento e teste são definidas na Tabela 21.

Tabela 21: Distribuição dos dados de treinamento e teste para cada faixa etária do estimador hierárquico por décadas (FG-NET).

Classe	Faixa etária	Número de amostras	Treinamento	Teste
1	0-9	371	278	93
2	10-19	339	254	85
3	20-29	144	108	36
4	30-39	79	61	18
5	40-49	46	34	12
6	50-59	15	11	4
7	60-69	8	6	2
Total		1.002	752	250

Os resultados da estimativa hierárquica por décadas em diferentes conjuntos de características são mostrados na Tabela 22. Para se ter um ideia do desempenho em cada uma das faixas etárias o *MAE* individual das décadas foram colocados na Tabela 22. A combinação de características globais e locais AAM+LPQ<sub>9×9</sub> foi a que obteve o melhor desempenho médio em termos de *MAE*, 5,98 anos. A taxa de acerto atingida na etapa da classificação em faixas etárias foi de 59,6%. O *MAE* das décadas 0-9 e 10-19 anos foram 4,82 e 6,61, respectivamente. A faixa etária (0-19 anos) contém a maior quantidade de dados para treinamento e teste, o qual justifica o *MAE* mais baixo, além dos indivíduos dessa faixa etária possuírem características mais relevantes com relação ao formato craniofacial. As décadas 20-29 e 30-39 tiveram um *MAE* mais elevado, 7,11 e 9,53,

respectivamente, devido a ter poucos dados de treinamento e teste e também devido os indivíduos dessa faixa etária possuírem características semelhantes como o formato facial e a textura da pele. As décadas 40-49, 50-59 e 60-69 são as que possuem a quantidade mais limitada de dados, o qual prejudica o desempenho do sistema. Pelo fato da década 60-69 possuir apenas 2 imagens para teste, os dados foram classificadas em outra classe durante a etapa de classificação, portanto não houve regressão na classe específica.

Em geral o desempenho da estimativa hierárquica por décadas é razoável, a taxa de acerto média atingida na etapa da classificação em faixas etárias foi de 51,61%, enquanto o resultado da etapa da estimativa específica utilizando regressores foi de 9,19 anos de *MAE*. O fato de possuir 7 classes com faixas menores de idade prejudica o sistema principalmente na etapa da classificação onde o erro é maior, pois algumas classes como a 6 e a 7 possuem poucos dados para treinamento e teste. Outro fator negativo é o número de classes e regressores os quais aumentam o custo computacional do sistema.

Tabela 22: Resultados da estimativa hierárquica por décadas em diferentes conjuntos de características (FG-NET).

Características	Classificação Taxa de acerto [%]	MAE anos	0-9 MAE	10-19 MAE	20-29 MAE	30-39 MAE	40-49 MAE	50-59 MAE	60-69 MAE
AAM	58,8	7,61	4,64	6,44	7,81	8,24	7,52	11,03	-
GW <sub>6,4</sub>	52,4	7,64	6,05	8,40	8,43	14,52	0,79	-	-
GW <sub>8,5</sub>	58,4	7,89	6,09	6,46	8,22	5,94	12,73	-	-
LBP <sub>8,1</sub>	40,8	9,06	8,92	10,45	7,79	-	-	-	-
LBP <sub>8,2</sub>	46	11,65	6,89	9,09	9,54	11,51	21,21	-	-
LBP <sub>8,3</sub>	49,2	8,10	6,69	8,00	10,09	6,70	9,03	-	-
LBP <sub>16,1</sub>	46	9,62	9,47	9,07	10,31	-	-	-	-
LBP <sub>16,2</sub>	48,4	11,02	5,92	8,31	8,31	10,48	22,06	-	-
LBP <sub>16,3</sub>	46	11,49	7,79	9,08	12,56	16,52	-	-	-
LPQ <sub>3×3</sub>	48,8	12,03	8,85	8,38	9,30	9,13	24,51	-	-
LPQ <sub>5×5</sub>	48	13,06	8,79	7,16	14,02	22,26	-	-	-
LPQ <sub>7×7</sub>	50	15,36	4,79	7,87	10,92	11,78	14,73	42,07	-
LPQ <sub>9×9</sub>	51,6	9,84	5,07	8,76	8,55	12,83	13,98	-	-
LPQ <sub>11×11</sub>	50,8	10,64	6,46	8,77	8,92	9,88	19,15	-	-
LBP <sub>(8,1)(8,2)</sub>	47,2	9,15	8,34	9,51	9,61	-	-	-	-
LBP <sub>(8,2)(8,3)</sub>	50,4	8,84	6,76	7,96	7,96	8,59	12,94	-	-
LBP <sub>(8,1)(8,3)</sub>	50,4	8,79	8,01	8,54	12,23	6,37	-	-	-
LBP <sub>(8,1)(8,2)(8,3)</sub>	50,4	11,79	7,65	8,33	8,55	9,35	25,06	-	-
LBP <sub>(16,1)(16,2)</sub>	50,4	10,11	6,69	7,63	9,70	13,15	13,37	-	-
LBP <sub>(16,2)(16,3)</sub>	47,6	7,59	7,45	9,25	9,03	8,30	3,92	-	-
LBP <sub>(16,1)(16,3)</sub>	48,8	8,53	7,33	9,99	8,26	-	-	-	-
LBP <sub>(16,1)(16,2)(16,3)</sub>	50	7,16	7,56	8,43	11,41	6,08	2,32	-	-
LPQ <sub>(5×5)(7×7)</sub>	51,6	9,04	5,76	7,32	11,90	12,55	16,68	-	-
LPQ <sub>(5×5)(9×9)</sub>	48,4	10,68	5,31	7,66	11,68	11,72	17,05	-	-
LPQ <sub>(5×5)(11×11)</sub>	50	12,01	6,16	8,06	11,55	15,88	18,42	-	-
LPQ <sub>(5×5)(7×7)(9×9)</sub>	48	9,00	5,97	6,70	12,83	7,39	12,10	-	-
LPQ <sub>(5×5)(7×7)(11×11)</sub>	50,4	9,28	5,60	7,90	12,91	6,38	13,59	-	-
AAM+GW <sub>6,4</sub>	59,2	7,21	3,27	6,99	8,14	6,58	11,07	-	-
AAM+GW <sub>8,5</sub>	60,4	6,51	2,90	7,30	5,79	10,14	8,94	3,98	-
AAM+LBP <sub>8,3</sub>	56,4	7,43	4,80	6,90	9,06	7,45	13,08	3,30	-
AAM+LBP <sub>16,3</sub>	50	8,89	5,73	8,18	10,49	6,78	13,26	-	-
AAM+LBP <sub>(8,2)(8,3)</sub>	50,4	9,37	5,84	8,60	8,94	13,36	16,93	2,55	-
AAM+LBP <sub>(16,1)(16,2)(16,3)</sub>	52,4	7,84	7,14	7,70	11,63	4,92	-	-	-
AAM+LPQ <sub>3×3</sub>	58,4	7,83	4,91	7,26	8,08	6,08	12,80	-	-
AAM+LPQ <sub>5×5</sub>	51,6	9,98	5,48	6,90	11,79	10,45	15,29	-	-
AAM+LPQ <sub>7×7</sub>	54,4	8,31	4,22	7,46	9,57	5,80	14,53	-	-
AAM+LPQ <sub>9×9</sub>	59,6	<b>5,98</b>	4,82	6,61	7,11	9,53	6,16	1,65	-
AAM+LPQ <sub>11×11</sub>	55,2	7,68	5,51	7,31	6,87	7,21	8,62	10,59	-
AAM+LPQ <sub>(5×5)(9×9)</sub>	53,6	8,38	5,82	6,97	8,95	10,29	9,85	-	-
AAM+GW <sub>8,5</sub> +LBP <sub>(16,1)(16,2)(16,3)</sub>	52,4	11,13	5,52	8,09	8,13	7,66	26,24	-	-
AAM+GW <sub>8,5</sub> +LPQ <sub>5×5</sub>	56,8	9,05	5,72	9,08	10,39	16,72	-	-	-
AAM+LBP <sub>(16,1)(16,2)(16,3)</sub> +LPQ <sub>5×5</sub>	52	8,12	7,82	8,24	8,58	8,74	-	-	-
AAM+GW <sub>8,5</sub> +LBP <sub>(16,1)(16,2)(16,3)</sub> +LPQ <sub>5×5</sub>	57,6	9,49	4,92	6,41	11,90	12,54	11,65	-	-
<b>Média geral</b>	<b>51,61</b>	<b>9,19</b>	<b>6,20</b>	<b>7,90</b>	<b>9,68</b>	<b>9,83</b>	<b>13,57</b>	<b>-</b>	<b>-</b>

## APÊNDICE D - Resultados FG-NET *Aging*

Tabela 23: Resultados de diferentes métodos de estimação etária em diferentes conjuntos de características, mostrando também a dimensão original das características e depois da redução da dimensionalidade utilizando PCA (FG-NET).

Características	Dimensão Orig./PCA	Classificação Taxa de acerto [%]	Regressão MAE	Single-level MAE	Hierárquico MAE	Hierárquico regiões sobrepostas fixas MAE	Hierárquico regiões sobrepostas flexíveis MAE
AAM	-/62	63,6	2,94	5,64	5,91	5,11	5,65
GW <sub>6,4</sub>	4999/237	60	3,17	7,62	7,32	7,01	7,45
GW <sub>8,5</sub>	8331/180	62,4	3,27	6,44	7,54	7,40	7,44
LBP <sub>8,1</sub>	649/177	53,6	3,97	8,13	9,90	9,80	10,05
LBP <sub>8,2</sub>	649/200	61,6	4,20	7,69	9,51	9,68	9,22
LBP <sub>8,3</sub>	649/163	61,6	3,39	6,98	8,45	8,28	8,18
LBP <sub>16,1</sub>	2673/171	60,8	3,22	7,79	8,95	9,14	8,83
LBP <sub>16,2</sub>	2673/152	64,4	3,41	7,63	6,84	7,72	6,42
LBP <sub>16,3</sub>	2673/241	60,4	3,66	7,11	6,42	6,23	6,12
LPQ <sub>3×3</sub>	2816/132	59,2	4,38	7,90	9,16	8,53	8,94
LPQ <sub>5×5</sub>	2816/200	62,8	3,72	6,98	6,60	6,49	6,26
LPQ <sub>7×7</sub>	2816/197	61,6	3,64	6,62	7,31	7,35	6,79
LPQ <sub>9×9</sub>	2816/195	60,4	3,51	6,74	6,35	5,82	5,74
LPQ <sub>11×11</sub>	2816/192	62,4	3,54	6,78	6,25	6,82	6,50
LBP <sub>(8,1)(8,2)</sub>	377/194	59,2	4,10	7,74	8,32	9,06	8,24
LBP <sub>(8,2)(8,3)</sub>	363/169	62,8	3,91	7,07	8,10	8,37	7,72
LBP <sub>(8,1)(8,3)</sub>	340/51	59,2	3,47	7,53	8,71	8,77	8,91
LBP <sub>(8,1)(8,2)(8,3)</sub>	540/246	60	3,56	7,16	8,50	8,18	7,64
LBP <sub>(16,1)(16,2)</sub>	323/159	63,6	3,14	7,51	8,97	9,40	8,92
LBP <sub>(16,2)(16,3)</sub>	393/176	63,2	3,65	7,27	6,84	7,11	6,38
LBP <sub>(16,1)(16,3)</sub>	412/247	60	3,64	7,49	6,39	6,54	6,29
LBP <sub>(16,1)(16,2)(16,3)</sub>	564/206	63,6	3,47	7,38	6,16	6,28	6,03
LPQ <sub>(5×5)(7×7)</sub>	397/213	63,6	3,40	6,56	6,81	6,89	6,43
LPQ <sub>(5×5)(9×9)</sub>	395/212	63,2	3,34	6,42	5,44	5,49	4,86
LPQ <sub>(5×5)(11×11)</sub>	392/204	63,2	3,46	6,55	7,08	6,66	6,60
LPQ <sub>(5×5)(7×7)(9×9)</sub>	592/164	64,4	3,71	6,46	6,34	6,13	6,05
LPQ <sub>(5×5)(7×7)(11×11)</sub>	589/186	65,2	3,74	6,52	7,09	7,12	6,68
AAM+GW <sub>6,4</sub>	299/54	66,4	<b>2,61</b>	5,49	6,02	5,98	6,06
AAM+GW <sub>8,5</sub>	242/58	63,6	2,86	<b>5,31</b>	5,61	5,54	5,25
AAM+LBP <sub>8,3</sub>	225/201	64,8	3,66	5,99	6,32	6,02	6,50
AAM+LBP <sub>16,3</sub>	303/191	63,2	3,54	6,25	5,92	6,25	5,32
AAM+LBP <sub>(8,2)(8,3)</sub>	231/261	65,6	3,44	6,68	6,24	6,62	6,03
AAM+LBP <sub>(16,1)(16,2)(16,3)</sub>	268/265	65,2	3,20	6,93	5,42	6,53	5,17
AAM+LPQ <sub>3×3</sub>	194/170	62	3,76	6,28	7,10	6,43	6,88
AAM+LPQ <sub>5×5</sub>	262/215	63,6	3,39	6,12	<b>5,13</b>	5,19	4,73
AAM+LPQ <sub>7×7</sub>	259/158	60	3,51	5,51	5,31	5,30	<b>4,50</b>
AAM+LPQ <sub>9×9</sub>	257/129	<b>68,8</b>	3,18	<b>5,31</b>	5,54	5,69	5,50
AAM+LPQ <sub>11×11</sub>	254/102	66,4	2,96	5,72	5,91	5,93	5,41
AAM+LPQ <sub>(5×5)(9×9)</sub>	274/222	67,2	3,23	6,08	5,51	5,31	4,79
AAM+GW <sub>8,5</sub> +LBP <sub>(16,1)(16,2)(16,3)</sub>	448/297	63,6	3,25	6,51	5,28	6,06	4,86
AAM+GW <sub>8,5</sub> +LPQ <sub>5×5</sub>	442/137	64,4	3,22	5,78	5,36	<b>4,85</b>	5,12
AAM+LBP <sub>(16,1)(16,2)(16,3)</sub> +LPQ <sub>5×5</sub>	468/329	65,2	3,30	6,38	5,33	4,96	4,95
AAM+GW <sub>8,5</sub> +LBP <sub>(16,1)(16,2)(16,3)</sub> +LPQ <sub>5×5</sub>	648/151	68	3,21	5,88	5,17	5,79	5,32
<b>Média geral</b>	-	62,9	3,46	6,70	6,80	6,83	<b>6,53</b>

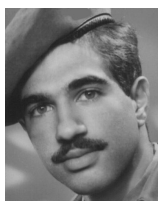
## APÊNDICE E – Resultados MORPH *Album 2*

Tabela 24: Resultados de diferentes métodos de estimação etária em diferentes conjuntos de características, mostrando também a dimensão após a redução da dimensionalidade utilizando PCA (MORPH).

Características	Dimensão PCA	Classificação Taxa de acerto [%]	Hierárquico <i>MAE</i>	Hierárquico regiões sobrepostas flexíveis <i>MAE</i>
AAM	107	31,31	11,61	10,84
GW <sub>8,5</sub>	189	42,36	8,02	7,26
LBP <sub>8,1</sub>	165	47,92	7,14	6,36
LBP <sub>8,2</sub>	223	48,57	7,07	6,38
LBP <sub>8,3</sub>	263	49,23	7,16	6,39
LBP <sub>16,1</sub>	199	47,48	7,10	6,31
LBP <sub>16,2</sub>	207	47,55	7,18	6,29
LBP <sub>16,3</sub>	103	47,99	7,20	6,31
LPQ <sub>5×5</sub>	176	50,77	<b>6,67</b>	5,90
LPQ <sub>7×7</sub>	250	50,69	6,72	<b>5,86</b>
LPQ <sub>9×9</sub>	239	<b>51,06</b>	6,74	5,91
LBP <sub>(16,1)(16,2)(16,3)</sub>	251	48,87	7,04	6,17
AAM+LPQ <sub>5×5</sub>	274	47,84	7,15	6,27
AAM+LPQ <sub>7×7</sub>	280	48,35	7,27	6,31
AAM+GW <sub>8,5</sub> +LBP <sub>(16,1)(16,2)(16,3)</sub>	308	47,62	7,19	6,49
AAM+GW <sub>8,5</sub> +LPQ <sub>5×5</sub>	300	48,50	6,87	6,29
AAM+LBP <sub>(16,1)(16,2)(16,3)</sub> +LPQ <sub>5×5</sub>	209	50,48	6,79	5,93
<b>Média geral</b>	N/A	47,45	7,35	<b>6,55</b>

## APÊNDICE F – Formulário de pesquisa

Qual é a Idade Aparente?



Enviar