



Pós-Graduação em Ciência da Computação

RECONHECIMENTO FACIAL EM IMAGENS DE BAIXA RESOLUÇÃO

Por

JOSÉ IVSON SOARES DA SILVA

Dissertação de Mestrado



Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
www.cin.ufpe.br/~posgraduacao

RECIFE
2015

JOSÉ IVSON SOARES DA SILVA

RECONHECIMENTO FACIAL EM IMAGENS
DE BAIXA RESOLUÇÃO

*ESTE TRABALHO FOI APRESENTADO À PÓS-GRADUAÇÃO EM
CIÊNCIA DA COMPUTAÇÃO DO CENTRO DE INFORMÁTICA DA
UNIVERSIDADE FEDERAL DE PERNAMBUCO COMO REQUISITO
PARCIAL PARA OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIA DA
COMPUTAÇÃO.*

ORIENTADOR: TSANG ING REN
CO-ORIENTADOR: GEORGE DARMITON DA CUNHA CAVALCANTI

RECIFE
2015

Catálogo na fonte
Bibliotecária Joana D'Arc Leão Salvador CRB4-532

S586r Silva, José Ivson Soares da.
Reconhecimento facial em imagens de baixa resolução / José Ivson Soares da Silva. – Recife: O Autor, 2015.
76 f.: fig., tab.

Orientador: Tsang Ing Ren.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIN, Ciência da Computação, 2015.
Inclui referências.

1. Sistema de reconhecimento de padrão. 2. Processamento de imagens. 3. Identificação biométrica. I. Tsang, Ing Ren (Orientador).
II. Título.

006.4 CDD (22. ed.) UFPE-MEI 2015-095

Dissertação de Mestrado apresentada por **José Ivson Soares da Silva** à Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, sob o título “**Reconhecimento Facial em Imagens de Baixa Resolução**”, orientada pelo **Prof. Tsang Ing Ren** aprovada pela Banca Examinadora formada pelos professores:

Profa. Teresa Bernarda Ludermir
Centro de Informática / UFPE

Prof. Hae Yong Kim
Departamento de Engenharia de Sistemas Eletrônicos / USP

Prof. Tsang Ing Ren
Centro de Informática / UFPE

Visto e permitida a impressão.
Recife, 24 de fevereiro de 2015.

Profa. Edna Natividade da Silva Barros
Coordenador da Pós-Graduação em Ciência da Computação do
Centro de Informática da Universidade Federal de Pernambuco.

*Este trabalho é dedicado aos meus pais
Marilene e Ivan.*

Agradecimentos

Gostaria de agradecer à minha família pelo apoio em todos os momentos dessa caminhada. Ao meu orientador Tsang pelas instruções ao longo desse trabalho. À equipe do laboratório OKI Brasil CIn UFPE pela disposição a ajudar quando preciso.

*"Fui moço e já, agora, sou velho,
porém jamais vi o justo desamparado,
nem a sua descendência a mendigar o pão."
(Salmos 37, 25)*

Resumo

Tem crescido o uso de sistemas computacionais para reconhecimento de pessoas por meio de dados biométricos, conseqüentemente os métodos para realizar o reconhecimento tem evoluído. A biometria usada no reconhecimento pode ser face, voz, impressão digital ou qualquer característica física capaz de distinguir as pessoas. Mudanças causadas por cirurgias, envelhecimento ou cicatrizes, podem não causar mudanças significativas nas características faciais tornando possível o reconhecimento após essas mudanças de aparência propositalis ou não. Por outro lado tais mudanças se tornam um desafio para sistemas de reconhecimento automático. Além das mudanças físicas há outros fatores na obtenção da imagem que influenciam o reconhecimento facial como resolução da imagem, posição da face em relação a câmera, iluminação do ambiente, oclusão, expressão. A distância que uma pessoa aparece na cena modifica a resolução da região da sua face, o objetivo de sistemas direcionados a esse contexto é que a influência da resolução nas taxas de reconhecimento seja minimizada. Uma pessoa mais distante da câmera tem sua face na imagem numa resolução menor que uma que esteja mais próxima. Sistemas de reconhecimento facial têm um menor desempenho ao tratar imagens faciais de baixa resolução. Uma das fases de um sistema de reconhecimento é a extração de características, que processa os dados de entrada e fornece um conjunto de informações mais representativas das imagens. Na fase de extração de características os padrões da base de dados de treinamento são recebidos numa mesma dimensão, ou seja, no caso de imagens numa mesma resolução. Caso as imagens disponíveis para o treinamento sejam de resoluções diferentes ou as imagens de teste sejam de resolução diferente do treinamento, faz-se necessário que na fase de pré-processamento haja um tratamento de resolução. O tratamento na resolução pode ser aplicando um aumento da resolução das imagens menores ou redução da resolução das imagens maiores. O aumento da resolução não garante um ganho de informação que possa melhorar o desempenho dos sistemas. Neste trabalho são desenvolvidos dois métodos executados na fase de extração de características realizada por *Eigenface*, os vetores de características são redimensionados para uma nova escala menor por meio de interpolação, semelhante ao que acontece no redimensionamento de imagens. No primeiro método, após a extração de características, os vetores de características e as imagens de treinamento são redimensionados. Então, as imagens de treinamento e teste são projetadas no espaço de características pelos vetores de dimensão reduzida. No segundo método, apenas os vetores de características são redimensionados e multiplicados por um fator de compensação. Então, as imagens de treinamento são projetadas pelos vetores originais e as imagens de teste são projetadas pelos vetores reduzidos para o mesmo espaço. Os métodos propostos foram testados em 4 bases de dados de reconhecimento facial com a presença de problemas de variação de iluminação, variação de expressão facial, presença óculos e posicionamento do rosto.

Palavras-chaves: Reconhecimento Facial. Baixa Resolução. Extração de Características.

Abstract

In the last decades the use of computational systems to recognize people by biometric data is increasing, consequently the efficacy of methods to perform recognition is improving. The biometry used for recognition can be face, voice, fingerprint or other physical feature that enables the distinction of different persons. Facial changes caused by surgery, aging or scars, does not necessarily causes significant changes in facial features. For a human it is possible recognize other person after these interventions of the appearance. On the other hand, these interventions become a challenge to computer recognition systems. Beyond the physical changes there are other factors in acquisition of an image that influence the face recognition such as the image resolution, position between face and camera, light from environment, occlusions and variation of facial expression. The distance that a person is at image acquisition changes the resolution of face image. The objective of systems for this context is to minimize the influence of the image resolution for the recognition. A person more distant from the camera has the image of the face in a smaller resolution than a person near the camera. Face recognition systems have a poor performance to analyse low resolution image. One of steps of a recognition system is the features extraction that processes the input data so provides more representative images. In the features extraction step the images from the training database are received at same dimension, in other words, to analyse the images they have the same resolution. If the training images have different resolutions of test images it is necessary a preprocessing to normalize the image resolution. The preprocessing of an image can be to increase the resolution of small images or to reduce the resolution of big images. The increase resolution does not guarantee that there is a information gain that can improves the performance of the recognition systems. In this work two methods are developed at features extraction step based on Eigenface. The feature vectors are resized to a smaller scale, similar to image resize. In first method, after the feature extraction step, the feature vectors and the training images are resized. Then the training and test images are projected to feature space by the resized feature vectors. In second method, only the feature vectors are resized and multiplied by a compensation factor. The training images are projected by original feature vectors and the test images are projected by resized feature vectors to the same space. The proposed methods were tested in 4 databases of face recognition with presence of light variation, variation of facial expression, use of glasses and face position.

Key-words: Face Recognition. Low Resolution. Feature Extraction.

Lista de Figuras

Figura 1 – Exemplos de biometrias. Voz, impressão digital, íris e face	18
Figura 2 – <i>Frames</i> de um vídeo gravado em resolução 640x480 pixels com a pessoa se aproximando da câmera.	22
Figura 3 – Diagrama de reconhecimento por verificação	22
Figura 4 – Diagrama de reconhecimento por identificação	23
Figura 5 – Diagrama de principais etapas de um sistema de reconhecimento facial	23
Figura 6 – Exemplo de características usadas na detecção facial	25
Figura 7 – Exemplo de imagem integral	26
Figura 8 – Exemplo de cascata de classificadores	26
Figura 9 – Esquema de <i>Mixture of Experts</i> (EBRAHIMPOUR et al., 2011)	33
Figura 10 – Diagrama de reconhecimento usando <i>Eigenface</i> com aumento das imagens de teste	36
Figura 11 – Diagrama de reconhecimento usando <i>Eigenface</i> com redução das imagens de treinamento	37
Figura 12 – Imagem média e quatro eigenvectors no tamanho original e reduzidos por interpolação bicúbica	38
Figura 13 – Diagrama de reconhecimento usando Redução de Características I . . .	39
Figura 14 – Diagrama de reconhecimento usando Redução de Características II . .	41
Figura 15 – Histograma com exemplo de limiar (linha vermelha) baseado nas distâncias do conjunto de validação	44
Figura 16 – Exemplo de imagens da base de dados ORL	46
Figura 17 – Exemplo de imagens da base de dados Yale	47
Figura 18 – Exemplo de imagens da base de dados Sheffield	47
Figura 19 – Exemplo de imagens da base de dados Georgia Tech	48
Figura 20 – Sensibilidade por resolução com imagens de treinamento, validação e teste da base de dados ORL	50
Figura 21 – Sensibilidade por resolução com imagens de treinamento, validação e teste da base de dados Yale	51
Figura 22 – Sensibilidade por resolução com imagens de treinamento, validação e teste da base de dados Sheffield	52
Figura 23 – Sensibilidade por resolução com imagens de treinamento das bases de dados ORL, Yale e Sheffield	53
Figura 24 – Especificidade por resolução com imagens de treinamento e validação da base de dados ORL e imagens de teste da base de dados Georgia Tech	54
Figura 25 – Especificidade por resolução com imagens de treinamento e validação da base de dados Yale e imagens de teste da base de dados Georgia Tech	55

Figura 26 – Especificidade por resolução por resolução com imagens de treinamento e validação da base de dados Sheffield e imagens de teste da base de dados Georgia Tech	56
Figura 27 – Especificidade por resolução com imagens de treinamento das bases de dados ORL, Yale e Sheffield e imagens de teste da base de dados Georgia Tech	57
Figura 28 – Precisão por resolução com imagens de treinamento e validação da base de dados ORL e imagens de teste das bases de dados ORL e Georgia Tech	59
Figura 29 – Taxa de precisão por resolução com imagens de treinamento e validação da base de dados Yale e imagens de teste das bases de dados Yale e Georgia Tech	59
Figura 30 – Precisão por resolução com imagens de treinamento e validação da base de dados Sheffield e imagens de teste das bases de dados Sheffield e Georgia Tech	60
Figura 31 – Precisão por resolução com imagens de treinamento e validação das bases de dados ORL, Yale e Sheffield e imagens de teste das bases de dados ORL, Yale, Sheffield e Georgia Tech	61
Figura 32 – Precisão Inversa por resolução com imagens de treinamento e validação da base de dados ORL e imagens de teste das bases de dados ORL e Georgia Tech	62
Figura 33 – Precisão Inversa por resolução com imagens de treinamento e validação da base de dados Yale e imagens de teste das bases de dados Yale e Georgia Tech	63
Figura 34 – Precisão Inversa por resolução com imagens de treinamento e validação da base de dados Sheffield e imagens de teste das bases de dados Sheffield e Georgia Tech	64
Figura 35 – Precisão Inversa por resolução com imagens de treinamento e validação das bases de dados ORL, Yale e Sheffield e imagens de teste das bases de dados ORL, Yale, Sheffield e Georgia Tech	65
Figura 36 – Acurácia por resolução com imagens de treinamento e validação da base de dados ORL e imagens de teste das bases de dados ORL e Georgia Tech	66
Figura 37 – Acurácia por resolução com imagens de treinamento e validação da base de dados Yale e imagens de teste das bases de dados Yale e Georgia Tech	66
Figura 38 – Acurácia por resolução com imagens de treinamento e validação da base de dados Sheffield e imagens de teste das bases de dados Sheffield e Georgia Tech	67
Figura 39 – Acurácia por resolução com imagens de treinamento e validação das bases de dados ORL, Yale e Sheffield e imagens de teste das bases de dados ORL, Yale, Sheffield e Georgia Tech	68

Figura 40 – Sensibilidade por resolução com imagens de treinamento das bases de dados ORL, Yale e Sheffield	68
Figura 41 – Especificidade por resolução com imagens de treinamento das bases de dados ORL, Yale e Sheffield e imagens de teste da base de dados Georgia Tech	69
Figura 42 – Precisão por resolução com imagens de treinamento e validação das bases de dados ORL, Yale e Sheffield e imagens de teste das bases de dados ORL, Yale, Sheffield e Georgia Tech	69
Figura 43 – Precisão Inversa por resolução com imagens de treinamento e validação das bases de dados ORL, Yale e Sheffield e imagens de teste das bases de dados ORL, Yale, Sheffield e Georgia Tech	70
Figura 44 – Acurácia por resolução com imagens de treinamento e validação das bases de dados ORL, Yale e Sheffield e imagens de teste das bases de dados ORL, Yale, Sheffield e Georgia Tech	70

Lista de Tabelas

Tabela 1	– Características presentes nas bases de dados usadas presentes no treinamento	48
Tabela 2	– Número de imagens no conjunto de treinamento, validação e teste . . .	48
Tabela 3	– Representação das possibilidades de classe real e classe estimada	49
Tabela 4	– Sensibilidade por resolução com imagens de treinamento, validação e teste da base de dados ORL. Média μ e desvio padrão σ em %.	50
Tabela 5	– Sensibilidade por resolução com imagens de treinamento, validação e teste da base de dados Yale. Média μ e desvio padrão σ em %.	51
Tabela 6	– Sensibilidade por resolução com imagens de treinamento, validação e teste da base de dados Sheffield. Média μ e desvio padrão σ em %. . .	52
Tabela 7	– Sensibilidade por resolução com imagens de treinamento das bases de dados ORL, Yale e Sheffield. Média μ e desvio padrão σ em %.	53
Tabela 8	– Especificidade por resolução com imagens de treinamento e validação da base de dados ORL e imagens de teste da base de dados Georgia Tech. Média μ e desvio padrão σ em %.	54
Tabela 9	– Especificidade por resolução com imagens de treinamento e validação da base de dados Yale e imagens de teste da base de dados Georgia Tech. Média μ e desvio padrão σ em %.	55
Tabela 10	– Especificidade por resolução com imagens de treinamento e validação da base de dados Sheffield e imagens de teste da base de dados Georgia Tech. Média μ e desvio padrão σ em %.	56
Tabela 11	– Especificidade por resolução com imagens de treinamento das bases de dados ORL, Yale e Sheffield. Média μ e desvio padrão σ em %.	57
Tabela 12	– Tempo médio de processamento do treinamento, validação e teste da base de dados ORL. Média μ e desvio padrão σ em segundos.	61
Tabela 13	– Tempo médio de processamento do treinamento, validação e teste da base de dados Yale. Média μ e desvio padrão σ em segundos.	62
Tabela 14	– Tempo médio de processamento do treinamento, validação e teste da base de dados Sheffield. Média μ e desvio padrão σ em segundos. . . .	63
Tabela 15	– Tempo médio de processamento do treinamento, validação e teste da base de dados ORL, Yale e Sheffield. Média μ e desvio padrão σ em segundos.	64

Siglas e Acrônimos

SVM	Support Vector Machine
KNN	K Nearest Neighbor
PCA	Principal Components Analysis
2DPCA	Two Dimensional Principal Component Analysis
IMPCA	Image Principal Component Analysis
HFIF	Hallucinating Facial Images and Features
SVDD	Support Vector Data Description
SIFT	Scale-Invariant Feature Transform
SURF	Speeded Up Robust Features
LDA	Linear Discriminant Analysis
MLP	Multilayer Perceptron
DCT	Discrete Cosine Transform
FFT	Fast Fourier Transform

Sumário

1	INTRODUÇÃO	17
1.1	Biometria	17
1.1.1	Voz	18
1.1.2	Impressão digital	19
1.1.3	Íris	19
1.2	Biometria Facial	19
1.2.1	Histórico	20
1.2.2	Desafios	21
1.3	Identificação X Verificação	21
1.4	Arquitetura de um sistema de reconhecimento	22
1.5	Objetivo	24
1.6	Organização do documento	24
2	SISTEMAS DE RECONHECIMENTO FACIAL	25
2.1	Deteção facial	25
2.2	Pré-processamento	27
2.2.1	Interpolação bicúbica	27
2.3	Extração de características	27
2.3.1	<i>Principal Component Analysis</i>	28
2.3.2	<i>Eigenface</i>	31
2.4	Classificação	31
2.4.1	KNN	33
3	MÉTODOS PROPOSTOS	35
3.1	Redução de Características I	37
3.2	Redução de Características II	38
3.3	Determinação de limiar	41
4	EXPERIMENTOS E RESULTADOS	45
4.1	Bases de dados	46
4.1.1	ORL	46
4.1.2	Yale	46
4.1.3	Sheffield (UMIST)	46
4.1.4	Georgia Tech Face Database	47
4.2	Protocolo de treinamento e teste	47
4.3	Resultados	48

4.3.1	Sensibilidade	49
4.3.2	Especificidade	52
4.3.3	Precisão	58
4.3.4	Precisão Inversa	58
4.3.5	Acurácia	60
4.3.6	Tempo de processamento	60
4.4	Experimentos adicionais	62
4.4.1	Interpolação bilinear	62
4.5	Considerações	64
	5 CONCLUSÃO	71
5.1	Trabalhos futuros	72
	Referências	73

Capítulo 1

Introdução

Os seres humanos usam características físicas ou comportamentais, como voz, face ou altura, para reconhecer outros indivíduos. Nas interações do cotidiano reconhecemos pessoas através de uma ou mais características com uma boa eficácia. Mudanças causadas por cirurgias, envelhecimento ou cicatrizes, podem não causar mudanças significativas nas características faciais tornando possível o reconhecimento após essas mudanças de aparência propositais ou não. Por outro lado tais mudanças se tornam um desafio para sistemas de reconhecimento automático. Pode-se dizer que a face seja a principal forma de identificação entre pessoas. Identificar pessoas através de sistemas computacionais se tornou uma boa opção para o controle automático de entrada de ambientes restritos, transações comerciais, registro de atividades e vigilância.

1.1 Biometria

Biometria é definida como reconhecimento automático de uma pessoa usando traços físicos ou comportamentais distinguíveis (DELAC e GRGIC, 2004). A identificação biométrica usa algumas características físicas e comportamentais únicas, não compartilhadas por qualquer outro indivíduo, para positivamente identificar um indivíduo. Muitas vezes usa-se biometria também para verificar se um indivíduo não está numa lista de procurados e não deseja ser identificado. A informação biométrica usada para o reconhecimento deve garantir que uma pessoa cadastrada seja reconhecida como ela mesma e que outras não sejam reconhecidas incorretamente. Desde a antiguidade as informações físicas eram usadas na identificação de pessoas, mesmo sem sistemas automatizados para realizá-la. A complexidade e eficácia da identificação biométrica vêm aumentando, mas os princípios de tentar extrair características únicas se mantêm. A Figura 1 apresenta alguns exemplos de biometrias e a seguir são descritos essas biometrias.

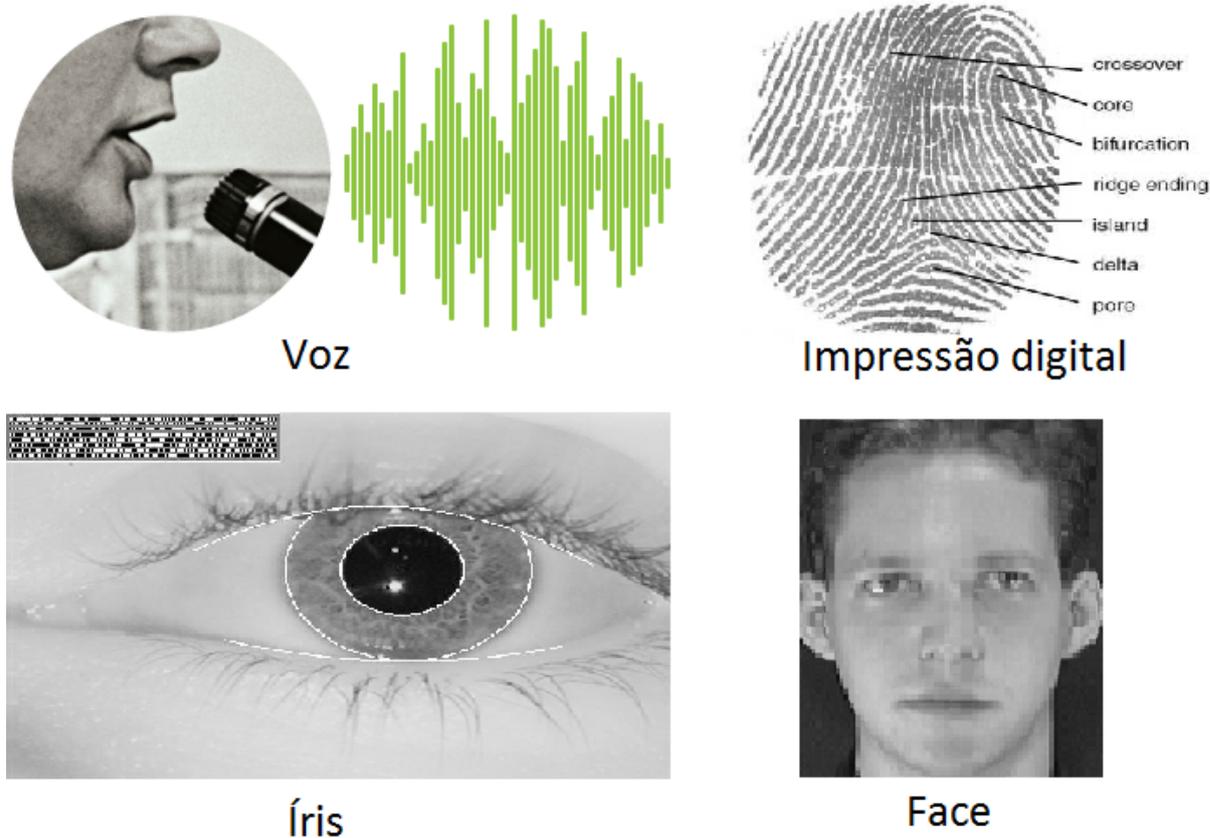


Figura 1 – Exemplos de biometrias. Voz, impressão digital, íris e face

1.1.1 Voz

As características da voz têm origem física e comportamental. Quanto a característica física, o trato vocal é o responsável pela produção da voz envolvendo laringe, faringe oral, cavidade oral, faringe nasal e cavidade nasal (CAMPBELL J.P., 1997). A voz pode ser usada em dois contextos, reconhecimento de fala e reconhecimento de locutor (ERTAS, 2011). No reconhecimento de locutor o objetivo é identificar quem está falando pela extração de informações do texto pronunciado. No reconhecimento de fala o objetivo é entender o que está sendo dito. Em ambos tipos de reconhecimento a extração de características é muito importante para comparação dos padrões e as medidas de distância são comuns em ambas aplicações. O reconhecimento de locutor é dividido em dois tipos, Reconhecimento de Locutor Dependente de Texto e Reconhecimento de Locutor Independente de Texto (REYNOLDS, 2002). No Reconhecimento de Locutor Dependente de Texto requer que o locutor pronuncie a mesma frase de cadastro e teste. No Reconhecimento de Locutor Independente de Texto, o locutor não é obrigado a pronunciar no momento do teste, a mesma frase pronunciada no cadastro.

1.1.2 Impressão digital

As impressões digitais de um indivíduo são formadas por cristas e vales da epiderme e têm características diferentes em cada dedo (JAIN et al., 1997). Os tipos de configuração são individualmente variáveis, mas variam dentro de um limite permitindo sua classificação sistemática. As configurações e detalhes de minúcias das cristas e vales individuais são permanentes e não sofrem alterações para um determinado dedo. Os sistemas de impressão digital possuem um alto nível de confiança com taxas de erro de identificação de 0.001%.

1.1.3 Íris

A íris é um órgão interno do olho que está bem protegido e tem uma estrutura estável ao longo do tempo (DAUGMAN, 2004). Há uma grande variabilidade de padrões da íris entre pessoas, esses padrões complexos são formados por ligamentos, sulcos, cristas, criptas, anéis, corona, sardas e colarinho em zigue-zague. A localização no rosto e a forma anelar facilitam o isolamento dessa característica e criação de uma representação de tamanho invariante. As informações extraídas da íris são insensíveis ao ângulo de iluminação. A íris possui uma alta capacidade de distinção entre duas pessoas, que garante a unicidade (DELAC e GRGIC, 2004). Os sistemas de reconhecimento de íris são bastante eficientes apresentando uma baixa taxa de falsas aceitações em grandes bases de dados. A implementação de sistemas de reconhecimento de íris tem um alto custo, é necessário proximidade entre o usuário e o equipamento de leitura ou o equipamento ser de alta qualidade na captura da imagem.

1.2 Biometria Facial

Reconhecimento facial é a principal maneira como as pessoas reconhecem umas as outras, embora outros sentidos também possam ser usados para esta tarefa como a voz. Apesar de o cérebro humano ter uma limitação no número total de pessoas que ele pode lembrar precisamente (ZHAO et al., 2003), a tarefa de reconhecimento é fácil para pessoas do seu convívio social mesmo em situações consideradas difíceis para um reconhecimento por sistema computacional. O reconhecimento facial pode ser usado em áreas de segurança para identificação de pessoas em multidões, já que é uma técnica não invasiva podendo ser usada sem a concordância do usuário e a distância seja pelo uso de câmeras de alta resolução de imagem ou métodos que sejam capazes de executar o reconhecimento com a imagem da face numa baixa resolução de imagem. Também pode ser usada na autenticação de identidade, no controle de acesso a ambientes, na vigilância de pessoas dentro de ambientes controlados por câmeras de vídeo. O problema de reconhecimento facial tem atraído atenção considerável de várias disciplinas e testemunhou um notável crescimento

nas pesquisas básicas e aplicadas, desenvolvimento de produto e aplicações (CHELLAPPA; SINHA; PHILLIPS, 2010).

1.2.1 Histórico

No final do século XIX o francês Alphonse Bertillon propôs um método para identificar indivíduos pelas medidas do corpo humano (BERTILLON, 1896). Bertillon se baseava na ideia de que as medidas do corpo de uma pessoa adulta não mudavam com o tempo, assim ele estava tentando resolver o problema de identificação de criminosos que forneciam nomes diferentes (MORAES, 2006). Uma das características usadas na identificação dos indivíduos era a face que naquele momento fora descrita de maneira verbal.

O primeiro sistema semi-automatizado para reconhecimento de face surgiu na década de 1960 (TAYLOR, 1967), neste sistema imagens de 10 pessoas eram apresentadas iterativamente e se o sistema respondesse de forma incorreta, então recebia um retorno com o nome da pessoa a ser identificada. O sistema construía uma matriz de pesos que eram atualizados a cada apresentação de uma nova imagem, semelhante ao funcionamento de uma rede neural. As mesmas faces eram usadas para treinamento e teste. Ao final da apresentação de cada imagem 250 vezes, então o sistema acertou a classificação de todas imagens apresentadas.

Na década de 1970 foi desenvolvido um sistema que utilizava 22 características para automatizarem o reconhecimento (GOLDSTEIN; HARMON; LESK, 1971). Algumas das características são sombra dos olhos, abertura do olho, ponta do nariz e largura da boca. Um dos objetivos deste trabalho era investigar o processo de reconhecimento feito por humanos. Os resultados do modelo proposto foram comparados com o reconhecimento feito por humanos. O modelo compara as características extraídas da imagem de teste com as das imagens de treinamento através de uma ordem de características e limiares. Um dos experimentos que utilizava 255 faces atingiu cerca de 70% de precisão no reconhecimento. KAYA e KOBAYASHI (1972) propuseram um trabalho semelhante em que também usaram características, mas usando distância Euclidiana entre os pontos marcados na face. Numa análise estatística, seria possível atingir 90% de precisão.

SAKAI; NAGAO; KANADE (1972) propuseram um trabalho semelhante aos dois anteriores, porém a extração de características era feita de forma automática a partir das imagens. Os melhores resultados foram obtidos com imagem da face completa e sem uso de óculos e barbas. KAUFMAN e BREEDING (1976) propuseram um sistema de reconhecimento de silhueta de perfil da face, o sistema atinge uma taxa de 90% para reconhecimento de 10 pessoas.

No trabalho de SIROVICH e KIRBY (1987) foi mostrado que eram necessárias

menos de uma centena de valores para representar uma face, para isso tais autores utilizaram o princípio da análise de componentes. Logo depois, TURK e PENTLAND (1991) propuseram o *Eigenface* e demonstraram que não eram necessários todos esses componentes para se realizar o reconhecimento, mas alguns resíduos poderiam ser importantes. Esta técnica consiste em projetar as imagens em novo espaço de características construído a partir de imagens de treinamento. As características selecionadas não representam necessariamente características físicas específicas.

A primeira modelagem 3D para reconhecimento da face foi proposta por GORDON (1991) e usava características de curvatura da face. As informações eram extraídas a partir de imagens 2D. A informação de profundidade torna invariante mudanças de aspectos como iluminação, ângulo e fundo, o que facilita o reconhecimento em situação, por exemplo, de oclusão parcial (LEO e MANIMEGALAI, 2011). São usadas duas abordagens para essa modelagem, a primeira consiste de um *scanner* que captura as posições da cabeça em 3D, a segunda reconstrói a face em 3D a partir de várias imagens 2D. O *Fisherface* (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) utiliza o *Eigenface* para reduzir a dimensão das imagens e então aplicar o LDA, o método apresentou um ganho de desempenho em bases de dados com variação de iluminação.

1.2.2 Desafios

O reconhecimento facial se torna mais difícil quando há variações na captura da imagem de um mesmo indivíduo, por exemplo, variações de iluminação, oclusão, pose, expressão e resolução. Com isso é importante que os métodos sejam robustos a essas variações e efetuem o reconhecimento mesmo na presença de tais dificuldades. O foco deste trabalho é o reconhecimento facial em imagens de diferentes resoluções. A Figura 2 mostra um problema de resolução, dependendo da distância entre a pessoa e a câmera a região recortada da face possui uma resolução diferente. Se o cadastro é realizado de uma forma controlada as imagens da região da face estarão numa boa resolução. Porém ao realizar um teste não se conhece a princípio qual a resolução da imagem de teste, imagens de face com menor resolução serão mais difíceis de identificar que aquelas numa resolução maior.

1.3 Identificação X Verificação

O reconhecimento biométrico se divide em dois grupos, identificação e verificação (CHELLAPPA; SINHA; PHILLIPS, 2010). Na verificação é fornecida a informação de quem o indivíduo se declara ser, então a comparação é realizada usando os dados biométricos fornecidos no momento da tentativa de reconhecimento e os dados biométricos cadastrados do indivíduo declarado, exemplo de diagrama na Figura 3.

Na identificação não é fornecida a informação de quem a pessoa declara ser, então



Figura 2 – *Frames* de um vídeo gravado em resolução 640x480 pixels com a pessoa se aproximando da câmera.

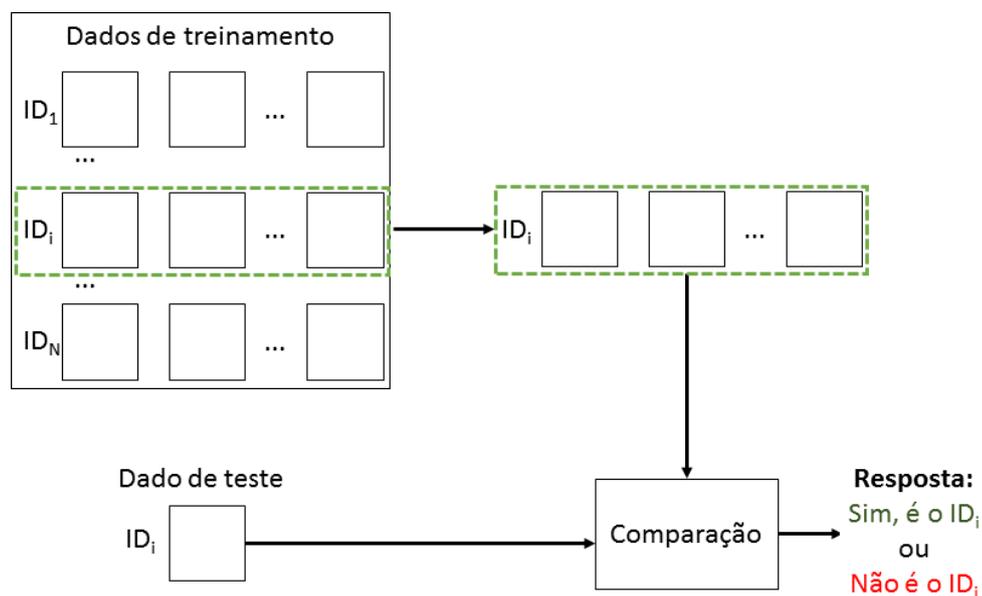


Figura 3 – Diagrama de reconhecimento por verificação

o sistema deve fazer uma comparação com dados de todos indivíduos cadastrados. Nos dois casos a decisão é baseada na similaridade entre os dados biométricos fornecidos e os recuperados, por exemplo, se a classificação for realizada por medidas de distância há um limiar, exemplo de diagrama na Figura 4.

1.4 Arquitetura de um sistema de reconhecimento

As principais etapas do reconhecimento facial são detecção, pré-processamento, extração de características e classificação, um diagrama das etapas é apresentado na Figura 5. Na fase de detecção cada imagem é analisada e as regiões que contém uma face são disponibilizadas para a etapa de pré-processamento. A detecção é importante para remoção do fundo da imagem que pode prejudicar as etapas seguintes do reconhecimento, se as imagens disponíveis forem recortadas manualmente esta etapa pode ser omitida. Na fase

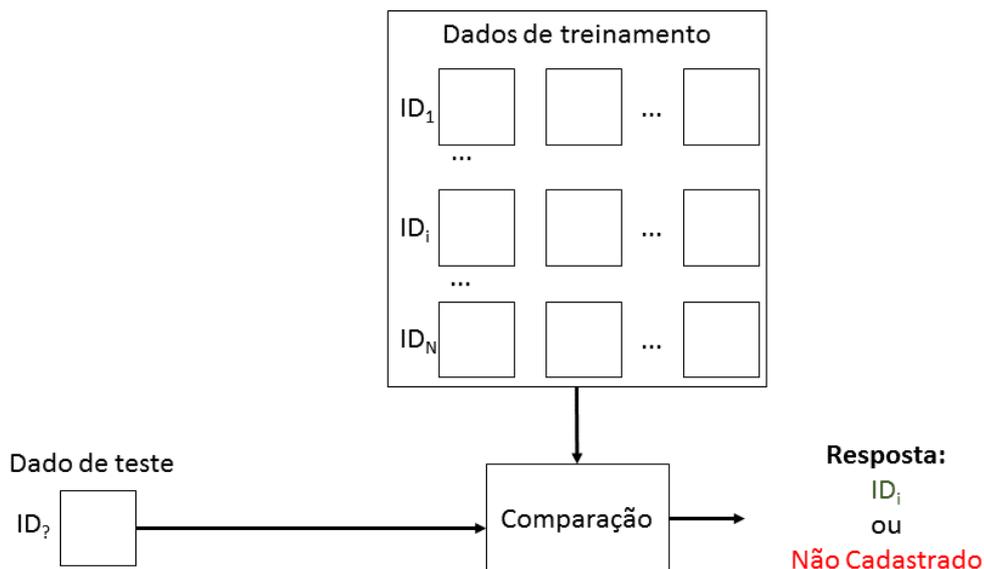


Figura 4 – Diagrama de reconhecimento por identificação

de pré-processamento são aplicados ajustes ou correções na imagem com o objetivo de reduzir ruído, redimensionamento das imagens, divisão das imagens por categorias. Na fase extração de características, são obtidas as informações mais importantes das imagens isso viabiliza uma classificação mais robusta. Na fase de classificação as características aprendidas durante o treinamento são comparadas com as características extraídas na fase de teste. A Figura 5 apresenta um digrama das etapas e as seções a seguir descrevem as etapas de forma mais detalhada.

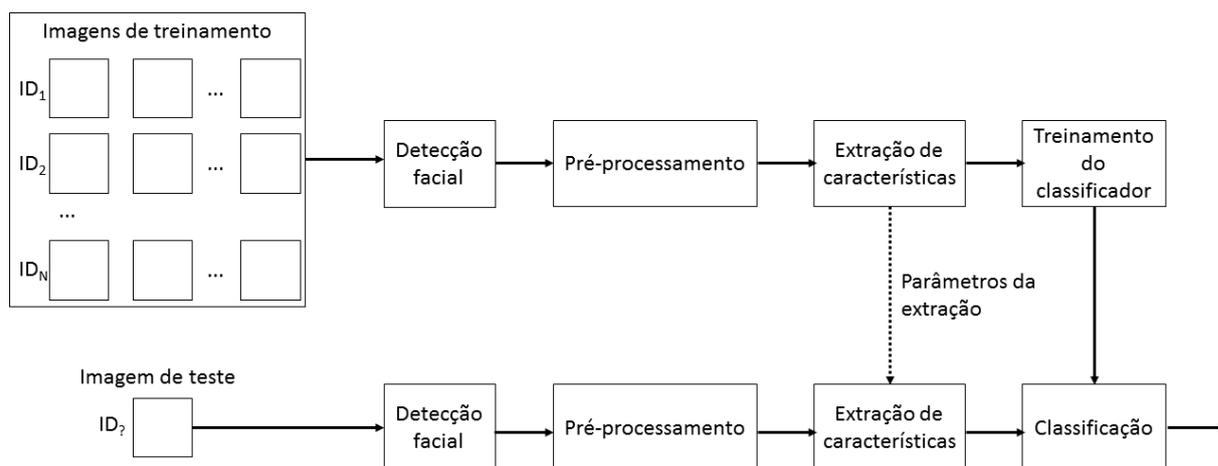


Figura 5 – Diagrama de principais etapas de um sistema de reconhecimento facial

1.5 Objetivo

O objetivo deste trabalho é propor métodos de reconhecimento facial que sejam capazes de realizar o reconhecimento facial em imagens de baixa resolução minimizando a queda de desempenho. A diferença na resolução entre as imagens faciais treinadas e testadas podem ser causadas por resoluções distintas de câmeras em cada fase ou pela distância entre a pessoa e câmera em cada fase. Nos experimentos as imagens de diferentes resoluções são simuladas aplicando redução interpolação nas imagens. A proposta é que a fase de extração de características seja realizada de maneira robusta, tal que a resolução da imagem não influencie esta etapa minimizando os erros na fase de classificação que poderiam ser causados pela resolução diferente das imagens. O método proposto consiste em reduzir os vetores de características obtidos na fase de treinamento e é baseado no *Eigenface*.

1.6 Organização do documento

O restante do documento é organizado da seguinte forma, o Capítulo 2 apresenta uma descrição das etapas de reconhecimento facial, detecção facial, pré-processamento, extração de características e classificação, e as técnicas usadas em cada etapa. O Capítulo 3 apresenta dois métodos propostos para o reconhecimento facial em imagens de baixa resolução através do redimensionamento de características extraídas do *Eigenface*. O Capítulo 4 apresenta os experimentos e resultados comparando os métodos propostos com métodos tradicionais. O Capítulo 5 apresenta uma conclusão do trabalho e trabalhos futuros.

Capítulo 2

Sistemas de Reconhecimento Facial

Neste capítulo são descritas as principais etapas de um sistema de reconhecimento facial. São apresentadas técnicas usadas no reconhecimento facial de baixa resolução ou que serão usadas no desenvolvimento do método proposto. Na etapa de detecção é apresentado o algoritmo de Detecção facial de VIOLA e JONES (2004), uma referência nessa área. Na etapa de pré-processamento é apresentada uma técnica é usada para aumentar a resolução das imagens nos experimentos, a interpolação bicúbica. Na etapa de extração de características é apresentado o PCA e o *Eigenface*, nos quais os métodos propostos são baseados. Por fim, na etapa de classificação é apresentado o KNN usado nos experimentos.

2.1 Detecção facial

A detecção facial é usada para identificar a área onde está a face na imagem. Um algoritmo bem difundido na área conhecido pelo nome dos autores VIOLA e JONES (2004) é baseado em 3 conceitos, as características retangulares exemplificadas na Figura 6, a imagem integral exemplificado na Figura 7 e a cascata de classificadores exemplificada na Figura 8.

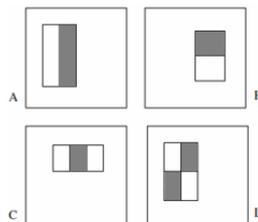


Figura 6 – Exemplo de características usadas na detecção facial

A detecção facial é realizada com base em características simples. A resolução base dessas características é 24×24 pixels, podendo ser elevadas para resoluções maiores

analisando a imagem em várias escalas. Cada característica é calculada subtraindo o somatório das regiões claras pelo somatório das regiões escuras, as regiões têm o mesmo tamanho e forma

$$c = \sum Regi\tilde{a}o_{clara} - \sum Regi\tilde{a}o_{escura}. \tag{2.1}$$

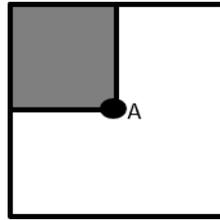


Figura 7 – Exemplo de imagem integral

Numa imagem integral cada pixel contém o valor da soma de todos os pixels da região acima e à esquerda. O uso da imagem integral acelera os cálculos das características retangulares, o valor do somatório de uma região é calculado a partir de 4 pontos de referência. Sendo x a imagem original, cada ponto da imagem integral é calculado da seguinte forma

$$I(i, j) = \sum_{i' \leq i, j' \leq j} x(i', j'). \tag{2.2}$$

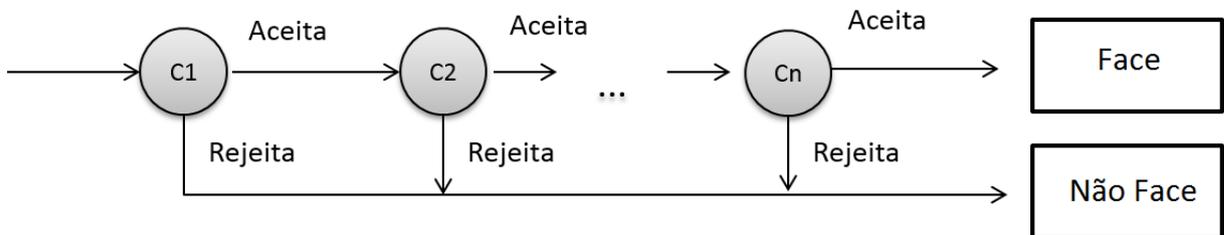


Figura 8 – Exemplo de cascata de classificadores

Vários classificadores fracos são utilizados em cascata para formar um classificador forte. Os classificadores são treinados usando AdaBoost (FREUND e SCHAPIRE, 1997), cada classificador fraco é treinado com um número pequeno de características. Se a região analisada na imagem for positivamente classificada por todos os classificadores da cascata, a região é classificada como face. Caso algum classificador da cascata classifique negativamente a imagem, a região é rejeitada como face. Um classificador fraco $h(x, f, p, \theta)$ consiste de uma característica (f), um *threshold* (θ) e uma polaridade (p) indicando a direção da desigualdade

$$h(x, f, p, \theta) = \begin{cases} 1 & \text{se } pf(x) < p\theta \\ 0 & \text{caso contrário} \end{cases} \tag{2.3}$$

2.2 Pré-processamento

A etapa de pré-processamento aplica ajustes na imagem para que pequenas variações não influenciem no resultado da extração de características, o ajuste aplicado pode ser correção de iluminação, estimativa de pose ou redimensionamento. Este trabalho trata o problema de resolução das imagens, então as técnicas descritas para pré-processamento são de redimensionamento de imagens.

2.2.1 Interpolação bicúbica

A interpolação é uma ferramenta usada no processamento de imagens para tarefas de ampliação, redução, rotação e correções geométricas (GONZALEZ e WOODS, 2000). Na interpolação os dados reais ou conhecidos são usado para estimar os valores dos novos pontos. Como na interpolação bicúbica são utilizados os 16 vizinhos mais próximo de um ponto para estimar seu valor. O valor do ponto (x, y) é estimado pela Equação

$$v(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij}x(i)y(j) \quad (2.4)$$

onde os coeficientes a_{ij} são determinados a partir das 16 equações de 16 incógnitas que podem ser escritas utilizando os 16 vizinhos mais próximos do ponto (x, y) .

Um método mais complexo que utiliza a interpolação bicúbica como um primeiro passo é o *Face Hallucination* introduzido por BAKER e KANADE (2000) e consiste em reconstruir uma imagem facial em alta resolução a partir de uma imagem facial de entrada em baixa resolução usando uma base de dados de treinamento de imagens faciais de alta resolução. *Face Hallucination* se baseia em métodos de super-resolução (SCHULTZ e STEVENSON, 1996) (HARDIE; BARNARD; ARMSTRONG, 1997). Muitas técnicas aplicam *Face Hallucination* ou Super Resolução como uma forma de pré-processamento para o reconhecimento. A face possui um estrutura bem definida, olhos, boca e nariz, que pode ser visto como um subconjunto de objetos como um todo. O objetivo do *Face Hallucination* é gerar algoritmos que sejam específicos para faces que superem os algoritmos gerais de Super Resolução.

2.3 Extração de características

Extração de características consiste em transformar os dados de entrada em características, esse processo obtém as informações mais relevantes possibilitando uma redução dos dados. Uma aplicação da extração de características é na área de visão computacional, em vez de se analisar uma imagem completa com informações que são redundantes e não compactadas é preferível ter dados que representem essa imagem de forma mais eficiente. Ao executar um método de reconhecimento de padrões em problemas com imagens nem

todos os *pixels* são essenciais, os dados redundantes podem ser eliminados sem prejudicar o desempenho do método. Além disso, variações no objeto a ser reconhecido como translação, rotação e escala afetam a precisão dos classificadores. Assim, faz-se necessária a utilização de algoritmos que reduzam a dimensionalidade de forma eficiente proporcionando uma representação dos padrões de uma forma robusta a essas alterações.

Após a extração de características pode haver ou não uma seleção de características, que baseada em métricas de desempenho num conjunto de validação seleciona um grupo das características extraídas. Após a extração e seleção de características, as operações são realizadas num espaço M -dimensional tal que $M < N$, sendo N a dimensão inicial dos dados de entrada.

Os métodos propostos são executados na fase de extração de características e utilizam os vetores de características extraídos pelo *Eigenface* que é baseado no *Principal Component Analysis*. É usado o *Eigenface* por obter os mesmos resultados do *Principal Component Analysis* e executar de forma mais rápida. Por sua vez o *Principal Component Analysis* foi escolhido como base para este trabalho por ser um método bastante usado e explorado na área de reconhecimento facial. *Principal Component Analysis* tem sido usado na etapa de extração de características de diversos métodos recentes na literatura e mostrado uma eficácia notável.

2.3.1 *Principal Component Analysis*

Principal Component Analysis, PCA, é um método estatístico que aplica transformações lineares sobre as variáveis originais, obtendo um conjunto menor de novas variáveis não correlacionadas que representam a maioria das informações contidas no conjunto de variáveis inicial. Este conjunto menor pode ser usado de forma mais eficiente que o conjunto original, que continha dados correlacionados ou redundantes (DUNTEMAN, 1989). A transformação linear feita pelo PCA projeta dados de um espaço N -dimensional para um espaço M -dimensional, tal que $M < N$. As coordenadas desse novo espaço não tem correlação, ou seja, o espaço é ortogonal e estas coordenadas preservam uma maior variância entre os dados.

PCA consiste em calcular os autovetores e autovalores, ou *eigenvectors* e *eigenvalues*, da matriz de covariância do conjunto de dados. Os autovalores são ordenados de forma decrescente, tal que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$, onde λ_i corresponde a i -ésima componente principal. O autovetor \mathbf{e}_1 associado ao primeiro autovalor fornece a maior variância, o autovetor \mathbf{e}_2 associado ao segundo autovalor fornece a segunda maior variância e assim por diante. Dessa forma, é possível eliminar alguns autovetores que não têm muita relevância na diferenciação dos dados sem perdas significativas de dados.

Se o conjunto de treinamento for de imagens, primeiramente cada imagem é

transformada num vetor coluna pela concatenação de suas colunas. Sendo D o número de imagens e N o número de *pixels* da imagem

$$\mathbf{x}_i = [p_1 \cdots p_N]^T; \quad i = 1, \dots, D. \quad (2.5)$$

Assume-se que os dados de treinamento possuem média zero. Caso não sejam, é realizada uma centralização na média

$$\bar{\mathbf{x}} = \frac{1}{D} \sum_{i=1}^D \mathbf{x}_i \quad (2.6)$$

$$\mathbf{w}_i = \mathbf{x}_i - \bar{\mathbf{x}}. \quad (2.7)$$

A matriz de covariância é definida como

$$S_x = \text{cov}(W) = \frac{1}{D} \sum_{i=1}^D \mathbf{w}_i \mathbf{w}_i^T = WW^T. \quad (2.8)$$

Os autovalores e autovetores desejados são calculados a partir da matriz de covariância S_x (LI et al., 2008). O objetivo é encontrar um conjunto de autovetores que tem a maior projeção possível das imagens centralizadas. A matriz de covariância S_x é simétrica com N linhas e N colunas.

Os autovetores correspondentes a autovalores diferentes de zero da matriz de covariância produzem uma base ortonormal para o subespaço no qual mais dados de imagem podem ser representados com uma pequena quantidade de erro. Sendo A uma matriz em que cada coluna é um autovetor de S_x , ordenados de forma que a primeira coluna possua o autovetor \mathbf{e}_1 associado ao maior autovalor λ_1 e a última coluna possua o autovetor \mathbf{e}_N associado ao menor autovalor λ_N . A matriz A pode ser usada como uma matriz de transformação

$$Y = A^T W. \quad (2.9)$$

Como cada \mathbf{y}_i é uma combinação linear de \mathbf{w}_i a média das projeções será igual a zero, $\frac{1}{D} \sum_{i=1}^D \mathbf{y}_i = 0$. A matriz de covariância de Y é definida como

$$S_y = \frac{1}{D} \sum_{i=1}^D \mathbf{y}_i \mathbf{y}_i^T = \frac{1}{D} \sum_{i=1}^D A^T \mathbf{w}_i \mathbf{w}_i^T A = A^T S_x A. \quad (2.10)$$

S_y é uma matriz diagonal cujos elementos são os autovalores de A

$$S_y = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{bmatrix}. \quad (2.11)$$

O fato dos valores fora da diagonal principal da matriz de covariância S_y serem 0, indica que elementos dos vetores \mathbf{y}_i não estão correlacionados. Os dados \mathbf{x}_i foram projetados para um novo espaço que não há correlação entre eles.

Para redução de dimensionalidade, os autovetores são classificados em ordem decrescente de acordo com seus correspondentes autovalores. Assim os primeiros autovalores serão os que têm maior variância entre os dados podendo reduzir o número de características para a representação de dados de uma imagem descartando algumas combinações lineares. Dessa forma uma imagem facial pode ser projetada num espaço com um menor número de dimensões que a imagem original. Selecionando os d primeiros autovetores, a projeção é definida como

$$\mathbf{y}_i = \mathbf{A}_d^T(\mathbf{w}_i). \quad (2.12)$$

Um critério para a seleção do números de autovetores é a conservação de energia. A energia é definida como o somatório dos autovalores da matriz de covariância, a energia preservada pela seleção dos d primeiros autovetores é dada pela equação

$$E_d = \frac{\sum_{i=1}^d \lambda_i}{\sum_{j=1}^N \lambda_j}. \quad (2.13)$$

MAHMUD *et al.* (2014) propuseram a seleção dos autovetores por algoritmo genético. Após a extração de características são atribuídos pesos a cada autovetor e calculado o desempenho do reconhecimento para estes pesos.

IMPCA (YANG e YANG, 2002) e 2DPCA (YANG *et al.*, 2004) são métodos que utilizam a matriz da imagem em duas dimensões em vez de transformá-la em um vetor coluna. A ideia é obter os vetores de projeção na direção da linhas. Sendo X uma imagem de dimensão $m \times n$, \mathbf{e} um vetor coluna unitário n -dimensional, deseja-se projetar X em \mathbf{e} pela transformação linear:

$$\mathbf{y} = X\mathbf{e}. \quad (2.14)$$

Assim, \mathbf{y} será um vetor projetado m -dimensional que representa o vetor de características projetado da imagem X . WANG *et al.* (2008) propuseram um método de reconhecimento facial em imagens coloridas baseado no 2DPCA. A ideia é transformar os *pixels* num vetor e usar a informação de cor como segunda dimensão.

DiaPCA (ZHANG; ZHOU; CHEN, 2006) busca os vetores de projeção ótimos das imagens faciais diagonais e que a correlação nas linhas e colunas sejam mantidas. O DiaPCA pode ser combinado com o 2DPCA o que melhora sua performance. Tendo D imagens de treinamento $X_k(k = 1, 2, \dots, D)$ de dimensão $m \times n$, para cada X_k é criada uma imagem facial diagonal B_k .

O Kernel PCA é uma extensão não-linear do PCA que computa os *eigenvectors* num espaço de características de alta dimensão (SCHÖLKOPF; SMOLA; MÜLLER, 1997), que posteriormente foi introduzido como extrator de características faciais (KIM; JUNG; KIM, 2002).

Modular PCA (GOTTUMUKKAL e ASARI, 2004) divide a imagem facial em subimagens e aplica o PCA em cada uma delas. Dividindo a imagem em sub-regiões se tem a vantagem de uma análise local ressaltando características importantes que poderiam ser desprezadas numa análise global devido a grande dimensionalidade da imagem. O cálculo dos pesos de cada região da imagem facial do Modular PCA é otimizado por algoritmo genético (CAVALCANTI; REN; PEREIRA, 2013). Baseado em Modular PCA, as informações de textura e profundidade podem ser combinadas (PARVATHY; NAVEEN; MONI, 2014).

2.3.2 Eigenface

O *Eigenface* foi proposto por TURK e PENTLAND (TURK e PENTLAND, 1991) e a ideia é tratar os dados com uma matriz de covariância de menor dimensão, diminuindo o custo computacional. Quando o número de imagens disponíveis D é menor que o número de dimensões, muitas vezes $D \ll N$, haverá $M - 1$ autovetores independentes, os restantes terão os respectivos autovalores iguais a 0. Assim, é possível calcular os autovetores da matriz N -dimensional resolvendo os autovetores de uma matriz $D \times D$. Por exemplo, se houver 20 imagens de dimensão 100×100 são calculados os autovetores de uma matriz 20×20 no lugar de uma matriz 10000×10000 , o que torna o processamento mais rápido. Os autovetores e autovalores também são chamados de *eigenvectors* e *eigenvalues*.

Considere a matriz $\frac{1}{D}W^TW$ e seus autovetores \mathbf{v}_i tal que

$$\frac{1}{D}W^TW\mathbf{v}_i = \phi_i\mathbf{v}_i. \quad (2.15)$$

Multiplicando os dois lados por W

$$\frac{1}{D}WW^TW\mathbf{v}_i = \phi_iW\mathbf{v}_i, \quad (2.16)$$

onde Wv_i são os autovetores de $S_x = WW^T$. Os autovalores ϕ_i de W^TW são os mesmos λ_i de WW^T , isso quer dizer que todas as direções relevantes do PCA podem ser extraídas desta matriz. Deseja-se calcular \mathbf{v}_i a partir de \mathbf{e}_i , assumindo que \mathbf{e}_i é ortonormal garante-se o mesmo para \mathbf{v}_i pela transformação:

$$\mathbf{v}_i = \frac{1}{\sqrt{D\lambda_i}}W^T\mathbf{e}_i. \quad (2.17)$$

2.4 Classificação

Em reconhecimentos de padrões a tarefa fundamental é atribuir uma classe a um objeto desconhecido a partir de seus dados. A partir de um conjunto de objetos com classes conhecidas, chamado de conjunto de treinamento é possível formular uma hipótese sobre a classe do novo objeto. Esta tarefa é atribuída ao classificador, que a partir do conjunto

de treinamento estima a classe que um objeto desconhecido pertence a partir das suas características (MITCHELL, 1997).

O objetivo do classificador é obter uma função de separação entre as classes do problema, assim ao ser apresentado um padrão de teste, a função é aplicada e retorna a classe que esse padrão supostamente pertence. Assim, $y = f(X)$ onde $X = x_1, x_2, \dots, x_n$ é um vetor de características que representa um padrão, y é o rótulo de uma das classes do problema e f é a função de mapeamento do classificador (BEALE e JACKSON, 1990).

Se o classificador for linear então a função representa uma reta e será do tipo $f(x) = w^T x + w_0 = 0$, onde w é um vetor de pesos e w_0 é um limiar de separação. Essa reta tem o intuito de separar os padrões de classes diferentes, um padrão de teste será classificado de acordo com o lado da reta em que for mapeado. A definição da classe é feita da seguinte forma

$$ID(x) = \begin{cases} 1, & \text{se } f(x) \leq 0 \\ 2, & \text{se } f(x) > 0 \end{cases} \quad (2.18)$$

Um classificador linear pode não ser capaz de separar as classes, pois a distribuição espacial dos padrões pode ser bem mais complexa e não linearmente separável. Para tais problemas podem ser usados classificadores neurais que poderão construir uma separação mais complexa entre as classes.

Existem dois tipos de abordagem, classificadores com aprendizado supervisionado e não supervisionado. Na abordagem supervisionada, as classes dos padrões de treinamento são conhecidas. Assim os padrões do conjunto de treinamento são do tipo $(x_i, y_i)_{i=1}^n$, onde cada x_i é vetor de características e y_i o seu respectivo rótulo. Já na abordagem não supervisionada a informação de classe não é levada em consideração durante o treinamento, apenas os vetores de características $(x_i)_{i=1}^n$ são conhecidos pelo classificador.

Qualquer tipo de classificador pode ser utilizado na fase classificação após a extração de características. ANGGRAINI (2014) utiliza o SOM na etapa de classificação após aplicar o PCA na extração de características. INDUMATHI e PALANIVEL (2014) utilizam HMM e SVM após o PCA para um sistema de reconhecimento invariante a pose.

Os classificadores também podem ser combinados. Combinação de classificadores consiste em combinar as opiniões de um conjunto de classificadores de padrões, assim busca-se uma opinião que seja melhor do que as respostas individuais (KUNCHEVA, 2004). O que se deseja é que a precisão da combinação seja maior que o melhor dos classificadores individuais. EBRAHIMPOUR *et al.* (2010) utilizam combinação de classificadores para reconhecer as faces, a ideia é usar uma MLP para cada extrator PCA, DCT e FFT e combinar a saída da MLP por *Stacked Generalization* (TING; WITTEN, 1999). Em outro trabalho de EBRAHIMPOUR *et al.* (2011) também é utilizada combinação de classificadores só que *Mixture de Experts* (JACOBS et al., 1991) no lugar de *Stacked*

Generalization.

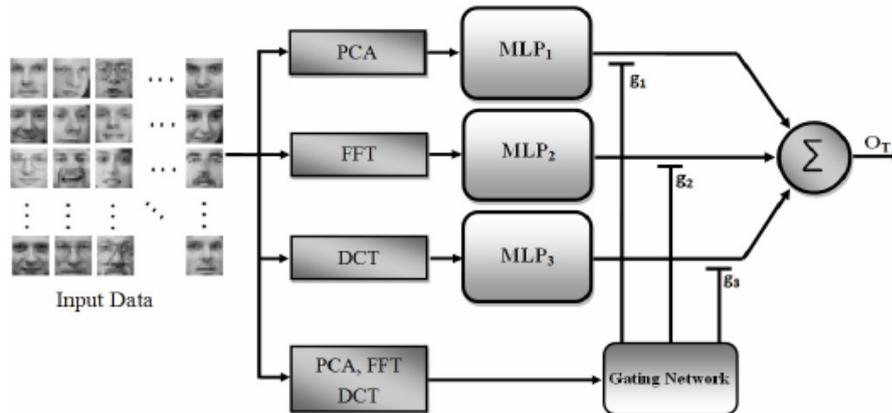


Figura 9 – Esquema de *Mixture of Experts* (EBRAHIMPOUR et al., 2011)

2.4.1 KNN

O *K-Nearest Neighbor* (KNN) é bastante usado em problemas de classificação, inclusive em reconhecimento facial (JOSE; POORNIMA; KUMAR, 2012). É um algoritmo abrangente que toma medidas de distâncias ou similaridades entre pares de padrões para realizar a classificação. Geralmente no caso de atributos numéricos a distância Euclidiana, Equação (2.20), entre dois padrões é usada (JIANG et al., 2007).

KNN armazena os exemplos de treinamento e seus respectivos rótulos de classes \mathbf{x}_i, c_i . O objetivo é classificar os exemplos de teste que tem o rótulo desconhecido, então dada uma instância \mathbf{x} se deseja obter o seu rótulo c . A determinação do rótulo é feita computando a distância entre o exemplo de teste e todos os exemplos de treinamento. Então são selecionados os k vizinhos mais próximos de \mathbf{x} , e o rótulo atribuído ao teste será o mais comum entre esses vizinhos como mostram o conjunto de Equações

$$\begin{cases} c(x) = \arg \max_{c \in C} \sum_{i=1}^k \delta(c, c(z_i)) \\ \delta(c, c(z_i)) = 1 \text{ se } c = c(z_i) \\ \delta(c, c(z_i)) = 0 \text{ caso contrário,} \end{cases} \quad (2.19)$$

onde z_1, z_2, \dots, z_k são os k vizinhos mais próximos de \mathbf{x} e C o conjunto de classes do problema.

Aplicando *Eigenface* em imagens faciais, primeiramente deve-se fazer uma projeção das imagens no novo espaço gerado pelo *Eigenface*. É esperado que imagens de faces do mesmo indivíduo sejam projetadas com uma pequena distância no espaço de características. Para cada uma das classes são obtidos os vetores de características de cada classe e a projeção das imagens no novo espaço. Durante a identificação de uma nova imagem o vetor de características dessa imagem é comparado com os vetores de características das

classes de pessoas. A classe que possuir a menor distância da imagem de teste é dita ser a classe a que o rosto pertence. A distância pode ser determinada pela Distância Euclidiana entre os vetores e a classificação via 1-NN, KNN com $k=1$.

$$DistanciaEuclidiana(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.20)$$

Capítulo 3

Métodos Propostos

O foco deste trabalho de mestrado é propor metodologias de reconhecimento de pessoas por face capaz de se executar o reconhecimento em imagens de baixa resolução minimizando a queda de desempenho. Este estudo tem aplicações práticas como em sistemas de segurança, onde as pessoas podem aparecer em diversos lugares da cena a distâncias que deixam a região da face na imagem analisada pequena.

Em métodos de extração de características como PCA as amostras de treinamento devem ter todas a mesma dimensionalidade assim como as amostras de teste. Quando uma imagem de teste tem a resolução diferente das imagens usadas no treinamento é necessário realizar um pré-processamento para que as imagens tenham uma mesma resolução. No capítulo anterior foram descritas técnicas para cada fase do reconhecimento, especialmente na fase de extração de características foram descritos o PCA e algumas variações. Os métodos propostos são baseados no *Eigenface* descrito na Seção 2.3.2. Os métodos propostos são comparados a duas abordagens implementadas com *Eigenface* (WANG; ZHANG; SHUM, 2004), a primeira consiste em aumentar a resolução da imagem de teste para a resolução em que foi realizado o treinamento que é chamada de *Eigenface* com aumento das imagens de teste, Figura 10. A segunda consiste em realizar o treinamento em diversas resoluções menores que a resolução original das imagens de treinamento que é chamada de *Eigenface* com redução das imagens de treinamento, Figura 11.

Quando o treinamento é realizado numa única resolução, a resolução original das imagens de treinamento, então para realizar a comparação de uma imagem de teste de resolução menor é necessário aplicar um pré-processamento como discutido na Seção 2.2. No caso de métodos mais simples como interpolação bicúbica não é garantida uma recuperação dos detalhes de regiões complexas das imagens facias como boca e olhos. Na interpolação bicúbica os novos dados são uma aproximação parametrizada pelas imagens de baixa resolução.

WANG; ZHANG; SHUM (2004) propuseram que a comparação seja realizada na

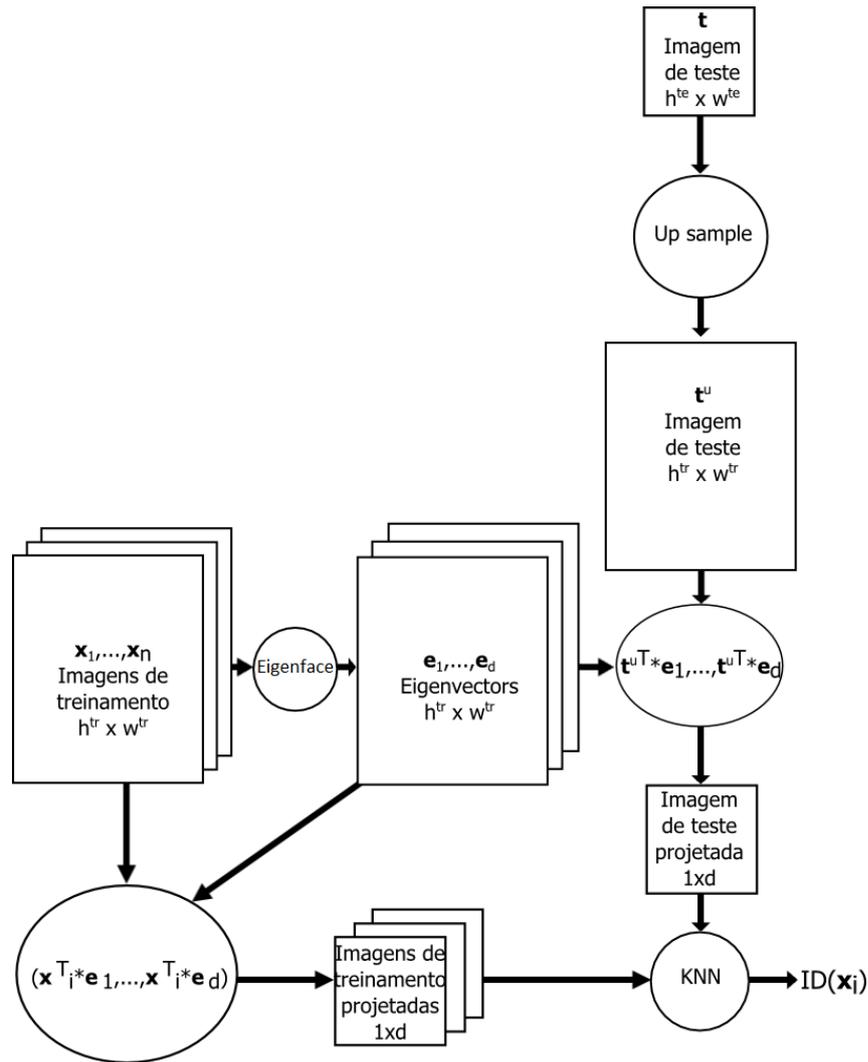


Figura 10 – Diagrama de reconhecimento usando *Eigenface* com aumento das imagens de teste

menor resolução conhecida. Para comparação com esse método as imagens de treinamento são redimensionadas para resoluções menores e o teste é realizado na resolução mais próxima. Essa abordagem segue o fluxo contrário do aumento do teste discutido anteriormente. Numa imagem de teste em baixa resolução o número de informações discriminantes ou detalhes é menor. Numa redução de escala por $\frac{1}{2}$, cada 4 *pixels* da imagem original serão reduzidos para 1 na imagem resultante. Reduzindo as imagens de treinamento as informações estruturais, comuns as imagens faciais, são preservadas. A maior perda de informações ocorre nos detalhes, as informações discriminantes que diferenciam as imagens faciais entre pessoas.

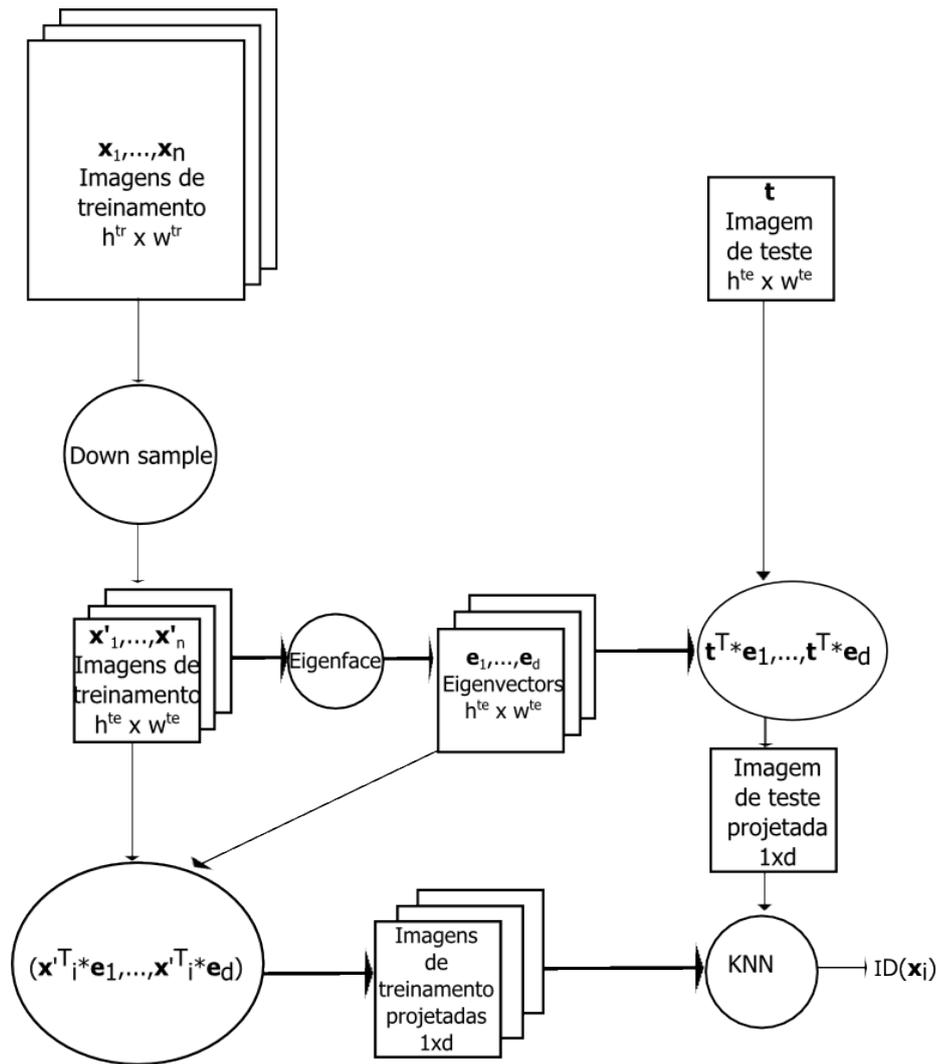


Figura 11 – Diagrama de reconhecimento usando *Eigenface* com redução das imagens de treinamento

3.1 Redução de Características I

Como já discutido anteriormente na Seção 2.3.2, os autovetores da matriz de covariância do *Eigenface*, chamados de *eigenvectors*, têm a mesma dimensão das imagens de treinamento. As imagens são processadas como vetores coluna e cada *eigenvector* também é um vetor coluna. Os *eigenvectors* podem ser remodelados para uma matriz de dimensão $h^{tr} \times w^{tr}$, sendo h^{tr} a altura e w^{tr} a largura das imagens de treinamento antes de serem transformadas num vetor coluna. A matriz do *eigenvector* pode ser analisada como uma imagem, a Figura 12 mostra a imagem média e alguns exemplos de *eigenvectors* em tons de cinza, normalizados para o intervalo $[0,255]$, extraídos da base de dados ORL descrita na Seção 4.1.1, e redimensionados por interpolação bicúbica. Apesar de perder a propriedade ortonormal dos autovetores, os vetores resultantes ainda podem projetar os dados de forma discriminante.



Figura 12 – Imagem média e quatro *eigenvectors* no tamanho original e reduzidos por interpolação bicúbica

Analisando os *eigenvectors* como imagens aplica-se uma interpolação bicúbica para reduzir a matriz para a dimensão da imagem de teste. Pela Figura 12 vemos que as características visuais da matriz do *eigenvector* se mantém e assim podemos supor que ainda é possível projetar uma imagem no espaço de características utilizando esses vetores. Redimensionando a matriz do *eigenvector* e a imagem de teste para vetores colunas pode-se então projetar o teste no espaço de características.

Em resumo, o treinamento é realizado com as imagens na resolução original, o *Eigenface* é aplicado sobre essas imagens, então as matrizes de transformação, *eigenvectors* e média, são salvas. As matrizes de transformação de características possuem vetores de dimensão compatível com o treinamento. Esses vetores são redimensionados para serem compatíveis com a dimensão das imagens de teste. As imagens de teste são projetadas no espaço de características usando a nova matriz de transformação. As imagens de treinamento são reduzidas para a dimensão de teste e também são projetadas no espaço de características. Uma vantagem em relação a redução das imagens de treinamento é que é necessário treinar apenas uma vez. A Figura 13 apresenta um diagrama de reconhecimento usando *Eigenface* quando o treinamento é realizado em resolução fixa e os *eigenvectors* são reduzidos para a dimensão do teste. O Algoritmo 1 apresenta os passos discutidos.

3.2 Redução de Características II

Se projetamos as imagens de treinamento usando os *eigenvectors* na sua dimensão original, as imagens projetadas de treinamento e a de teste têm a mesma dimensão, porém há uma diferença de escala. Como discutido anteriormente os vetores de características resultantes da redução de dimensão deixam de ser ortonormais. Experimentalmente foi verificado que a norma de cada vetor é reduzida pela mesma escala de redução aplicada a matriz do *eigenvector*, ou seja, numa redução de escala por $\frac{1}{2}$ a norma dos novos vetores torna-se aproximadamente 0,5.

Reescrevendo a Equação (2.9) com $A = [\mathbf{e}_1 \mathbf{e}_2 \cdots \mathbf{e}_d]$, onde \mathbf{e}_j para $j = 1, \dots, d$ é um

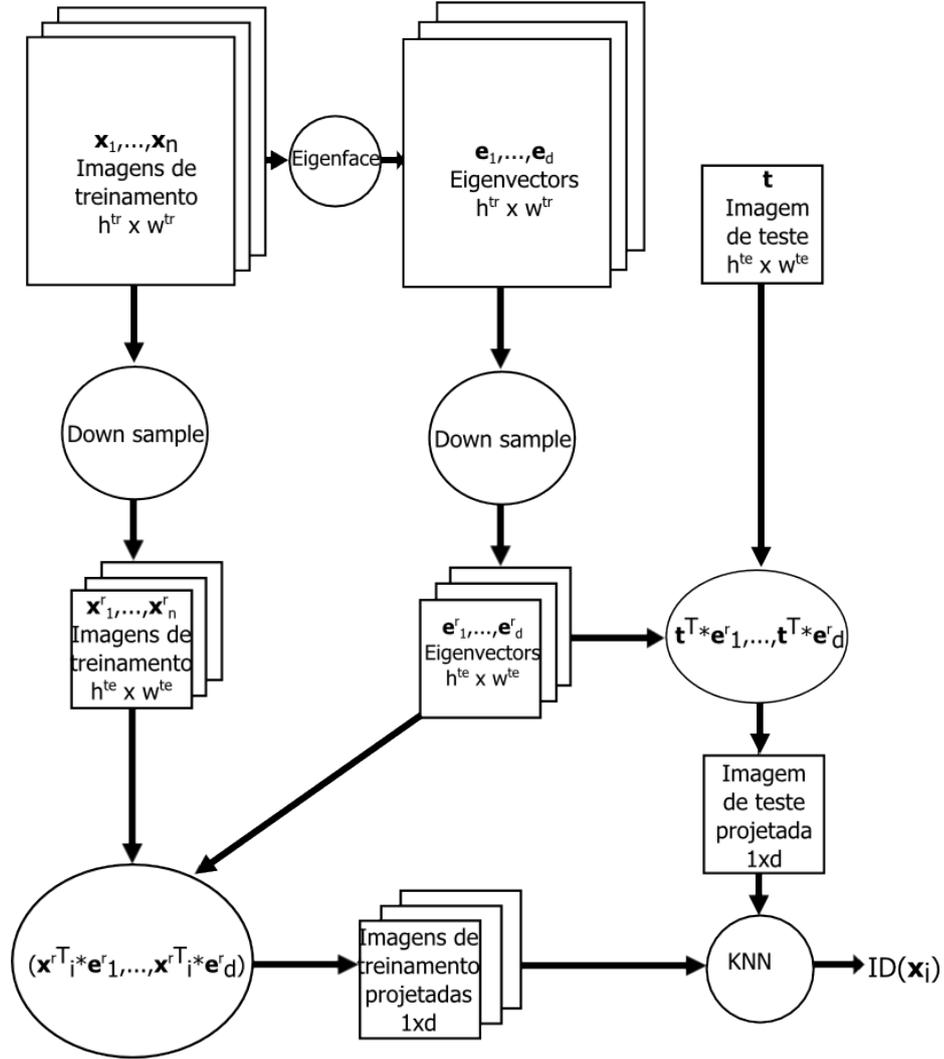


Figura 13 – Diagrama de reconhecimento usando Redução de Características I

vetor coluna correspondente ao j -ésimo *eigenvector*, e uma amostra \mathbf{x}_i centralizada na média das imagens de treinamento conforme as Equações (2.6) e (2.7)

$$\mathbf{y}_i = [\mathbf{e}_1 \mathbf{e}_2 \cdots \mathbf{e}_d]^T \mathbf{x}_i, \quad (3.1)$$

cada elemento \mathbf{y}_{ij} , $j = 1, \dots, d$ é dado por

$$\mathbf{y}_{ij} = \sum_{k=1}^N \mathbf{e}_{jk} \mathbf{x}_{ik}, \quad (3.2)$$

onde $N = h^{tr} \times w^{tr}$ é a dimensão de cada *eigenvector*. Quando redimensionados para $h^{te} \times w^{te}$, onde h^{te} a altura e w^{te} a largura da imagem de teste \mathbf{t} . O vetor coluna do *eigenvector* redimensionado \mathbf{e}_j^r passa a ter dimensão $M = h^{te} \times w^{te}$, a projeção da imagem de teste é dada por

$$\mathbf{z} = [\mathbf{e}_1^r \mathbf{e}_2^r \cdots \mathbf{e}_d^r]^T \mathbf{t}. \quad (3.3)$$

Algoritmo 1: Procedimento de teste com Redução de Características I

Entrada: Eigenvectors $E = [e_1, e_2, \dots, e_d]$
Entrada: Imagens de treinamento $X = [X_1, X_2, \dots, X_n]$
Entrada: Classe de imagens de treinamento $C = [C_1, C_2, \dots, C_n]$
Entrada: Imagem de teste Y
Saída: Classificação da imagem de teste
Saída: Distância da imagem de teste para a imagem de treinamento mais próxima

início

$h_{tr} \leftarrow \text{altura}(X_k);$
 $w_{tr} \leftarrow \text{largura}(X_k);$
 $h_{te} \leftarrow \text{altura}(Y);$
 $w_{te} \leftarrow \text{largura}(Y);$
 $\bar{X}(i, j) \leftarrow \frac{1}{n} \sum_{k=1}^n X_k(i, j),$ onde $i = 1, \dots, h_{tr}$ e $j = 1, \dots, w_{tr};$
 $temp_1 \leftarrow$ Redimensiona \bar{X} de $h_{tr} \times w_{tr}$ para $h_{te} \times w_{te};$
 $\bar{X}^{new} \leftarrow$ Remodela $temp_1$ de $h_{te} \times w_{te}$ para $1 \times (h_{te} \cdot w_{te});$
para $j=1$ até d **faça**
 $temp_1 \leftarrow$ Remodela vetor coluna e_j de $(h_{tr} \cdot w_{tr}) \times 1$ para $h_{tr} \times w_{tr};$
 $temp_2 \leftarrow$ Redimensiona $temp_1$ de $h_{tr} \times w_{tr}$ para $h_{te} \times w_{te};$
 $E_j^{new} \leftarrow$ Remodela $temp_2$ de $h_{te} \times w_{te}$ para $(h_{te} \cdot w_{te}) \times 1;$
fim
 $temp_1 \leftarrow$ Remodela Y para $1 \times (h_{te} \cdot w_{te});$
 $Y^r \leftarrow (temp_1 - \bar{X}^{new}) \cdot E^{new};$
para $k=1$ até n **faça**
 $temp_1 \leftarrow$ Redimensiona X_k de $h_{tr} \times w_{tr}$ para $h_{te} \times w_{te};$
 $temp_2 \leftarrow$ Remodela $temp_1$ de $h_{te} \times w_{te}$ para $1 \times (h_{te} \cdot w_{te});$
 $X_k^r \leftarrow (temp_2 - \bar{X}^{new}) \cdot E^{new};$
 $dist_k \leftarrow \text{DistanciaEuclidiana}(X_k^r, Y^r);$
fim
 $c \leftarrow \arg \min_k dist_k;$
 $Classificacao \leftarrow C_c;$
 $Distancia \leftarrow dist_c;$
retorna $[Classificacao, Distancia];$

fim

$$\mathbf{z}_j = \sum_{k=1}^M \mathbf{e}_{jk}^r \mathbf{t}_k. \quad (3.4)$$

Dado que $M < N$, o somatório da Equação (3.4) terá uma magnitude $\frac{N}{M}$ vezes menor que o da Equação (3.2), os dados serão projetados em regiões diferentes no espaço de características dificultando a classificação por medida de distância. Multiplicando os vetores reduzidos por $\frac{N}{M}$, os dados projetados ficam numa mesma escala. O método é resumido no diagrama da Figura 14 e Algoritmo 2.

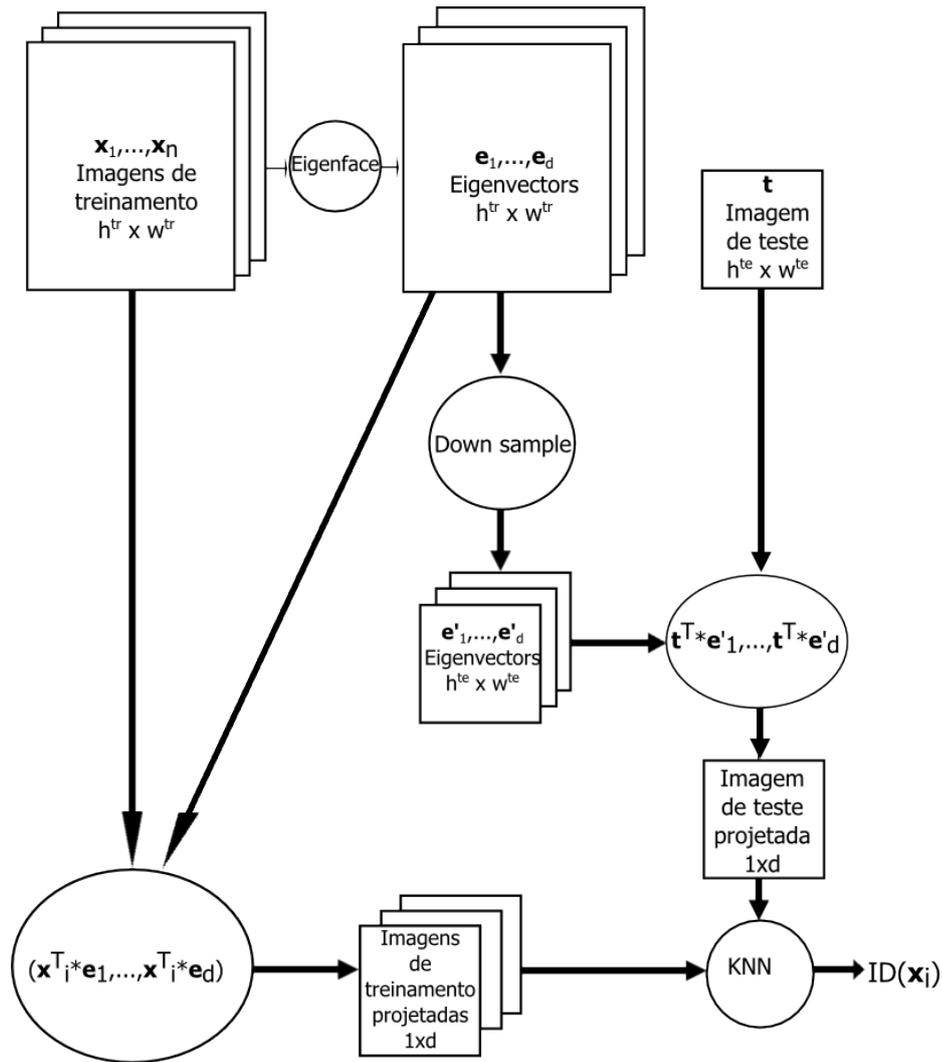


Figura 14 – Diagrama de reconhecimento usando Redução de Características II

3.3 Determinação de limiar

Quando o problema de reconhecimento é identificação geralmente a análise de desempenho é realizada com imagens de teste de indivíduos presentes na base de treinamento, então o erro possível é de estimar a classe errada sem levar em consideração a distância entre a imagem de teste e imagem de treinamento mais próxima. Se a análise for realizada apresentando imagens de teste de indivíduos que não estão na base de dados de treinamento, então é necessário analisar também a medida de distância. Para aceitar ou rejeitar a classe estimada é preciso estabelecer um limiar de separação da distância.

O conjunto de validação é usado para determinar o limiar de separação entre uma tentativa autêntica e uma tentativa de reconhecer alguém não cadastrado. As imagens de validação assim como as de teste possuem dimensões menores que as de treinamento. É aplicada a Distância Euclidiana entre as imagens de validação e as imagens de treinamento. São tomadas as distâncias das imagens de validação classificadas corretamente, as distâncias

Algoritmo 2: Procedimento de teste com Redução de Características II**Entrada:** Eigenvectors $E = [e_1, e_2, \dots, e_d]$ **Entrada:** Imagens de treinamento $X = [X_1, X_2, \dots, X_n]$ **Entrada:** Classe de imagens de treinamento $C = [C_1, C_2, \dots, C_n]$ **Entrada:** Imagem de teste Y **Saída:** Classificação da imagem de teste**Saída:** Distância da imagem de teste para a imagem de treinamento mais próxima**início** $h_{tr} \leftarrow \text{altura}(X_k);$ $w_{tr} \leftarrow \text{largura}(X_k);$ $h_{te} \leftarrow \text{altura}(Y);$ $w_{te} \leftarrow \text{largura}(Y);$ $\bar{X}(i, j) \leftarrow \frac{1}{n} \sum_{k=1}^n X_k(i, j),$ onde $i = 1, \dots, h_{tr}$ e $j = 1, \dots, w_{tr};$ $temp_1 \leftarrow$ Redimensiona \bar{X} de $h_{tr} \times w_{tr}$ para $h_{te} \times w_{te};$ $\bar{X}^{new} \leftarrow$ Remodela $temp_1$ de $h_{te} \times w_{te}$ para $1 \times (h_{te} \cdot w_{te});$ **para** $j=1$ até d **faça** $temp_1 \leftarrow$ Remodela vetor coluna e_j de $(h_{tr} \cdot w_{tr}) \times 1$ para $h_{tr} \times w_{tr};$ $temp_2 \leftarrow$ Redimensiona $temp_1$ de $h_{tr} \times w_{tr}$ para $h_{te} \times w_{te};$ $temp_3 \leftarrow temp_2 \cdot \frac{h_{tr} \cdot w_{tr}}{h_{te} \cdot w_{te}}$ $E_j^{new} \leftarrow$ Remodela $temp_3$ de $h_{te} \times w_{te}$ para $(h_{te} \cdot w_{te}) \times 1;$ **fim** $temp_1 \leftarrow$ Remodela Y para $1 \times (h_{te} \cdot w_{te});$ $Y^r \leftarrow (temp_1 - \bar{X}^{new}) \cdot E^{new};$ **para** $k=1$ até n **faça** $temp_1 \leftarrow$ Remodela X_k de $h_{tr} \times w_{tr}$ para $1 \times (h_{tr} \cdot w_{tr});$ $X_k^r \leftarrow (temp_1 - \bar{X}) \cdot E;$ $dist_k \leftarrow \text{DistanciaEuclidiana}(X_k^r, Y^r);$ **fim** $c \leftarrow \arg \min_k dist_k;$ $Classificacao \leftarrow C_c;$ $Distancia \leftarrow dist_c;$ **retorna** $[Classificacao, Distancia];$ **fim**

são normalizadas para o intervalo $[0,100]$ de maneira que a distância máxima de uma classificação correta seja 50 e mínima seja 0 na escala normalizada. O algoritmo 3 mostra os passos para obtenção do limiar.

Após ser aplicado o Algoritmo 1 ou o Algoritmo 2 para obter a classe e distância para o exemplo de treinamento mais próximo e o Algoritmo 3 para obter o limiar e os valores de máximo e mínimo podemos agora normalizar a distância das imagens de teste e validar a classe. Se a Distância Euclidiana normalizada do teste for menor que o limiar a imagem é aceita como autêntica, caso seja maior a imagem é dada como não reconhecida. Distância Euclidiana entre as imagens de teste e as imagens de treinamento.

Algoritmo 3: Procedimento de normalização das distâncias

Entrada: Eigenvectors $E = [e_1, e_2, \dots, e_d]$
Entrada: Imagens de treinamento $X = [X_1, X_2, \dots, X_n]$
Entrada: Classe de imagens de treinamento $C = [C_1, C_2, \dots, C_n]$
Entrada: Imagens de validação V_1, V_2, \dots, V_m
Entrada: Classe de imagens de validação D_1, D_2, \dots, D_m
Saída: Limiar
Saída: Máxima distância entre uma imagem de validação e uma imagem de treinamento da mesma classe
Saída: Mínima distância entre uma imagem de validação e uma imagem de treinamento da mesma classe

início

para $i=1$ até m **faça**

$[cVal_i, dVal_i] \leftarrow$ Aplique Procedimento de teste com Redução de Características(E, X, C, V_i);

fim

 Máximo $\leftarrow \max(dVal_i)$, onde $i = 1, \dots, m$ e $cVal_i = D_i$;
 Mínimo $\leftarrow \min(dVal_i)$, onde $i = 1, \dots, m$ e $cVal_i = D_i$;
 $dValNorm_i \leftarrow 100 \min(\max(dVal_i - \text{Mínimo}, 0) / (2(\text{Máximo} - \text{Mínimo}), 1))$; $i = 1, \dots, m$;
 $n_c \leftarrow |cVal_i = D_i|$, onde $i = 1, \dots, m$;

para $th=0$ até 100 **faça**

$n_t \leftarrow |dValNorm_i \leq th|$, onde $i = 1, \dots, m$ e $cVal_i = D_i$;
 se $n_t/n_c \geq 0.95$ **então**

 Limiar $\leftarrow th$;
 retorna [*Limiar, Máximo, Mínimo*];

fim

fim

Sendo I_{teste} a imagem de teste, C_{teste} a classificação da imagem de teste, $I_{proxima}$ a imagem de treinamento com menor distância para I_{teste} , $C_{proxima}$ a classe de $I_{proxima}$ e $d_{I_{teste}}$ a distância entre I_{teste} e $I_{proxima}$, a distância normalizada é dada pela Equação

$$d_{I_{teste}}^{norm} = \begin{cases} 100 \times (d_{I_{teste}} - \text{Mínimo}) / (2 \times (\text{Máximo} - \text{Mínimo})) & , \text{se } \text{Mínimo} < d_{I_{teste}} < \text{Máximo} \\ 0 & , \text{se } d_{I_{teste}} \leq \text{Mínimo} \\ 100 & , \text{se } d_{I_{teste}} \geq \text{Máximo} \end{cases} \quad (3.5)$$

Uma classificação de tentativa de autenticação de uma pessoa cadastrada é considerada correta se

$$d_{I_{teste}}^{norm} \leq \text{limiar} \text{ e } C_{proxima} = C_{teste}. \quad (3.6)$$

Uma classificação de tentativa de autenticação de impostor ou de imagem que não é face é considerada correta se

$$d_{I_{teste}}^{norm} > \text{limiar}. \quad (3.7)$$

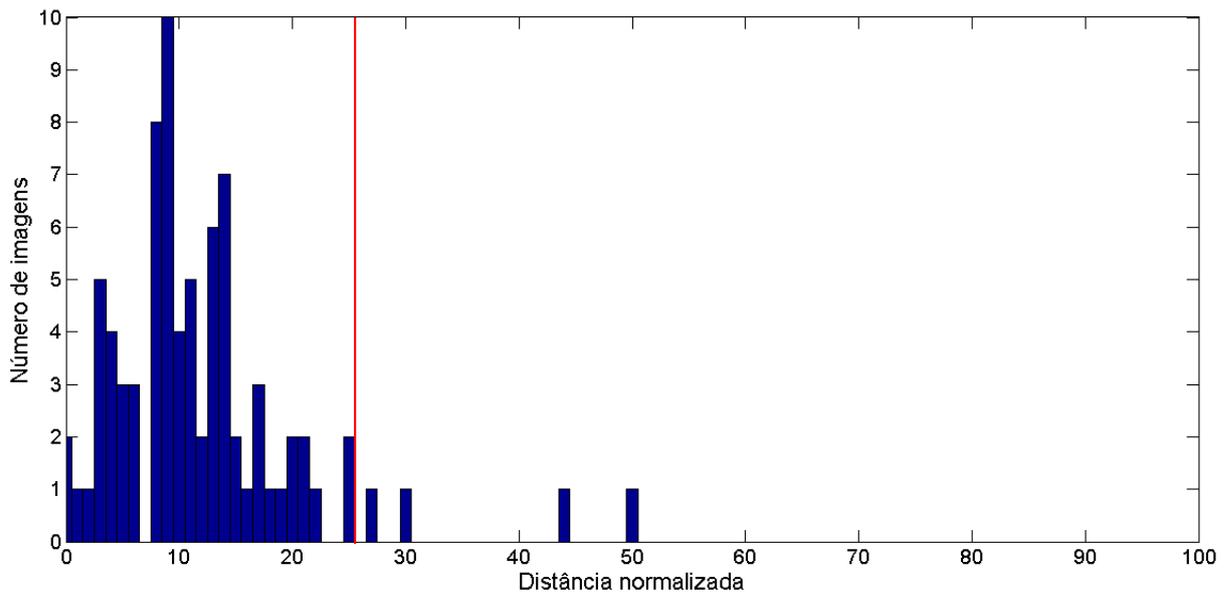


Figura 15 – Histograma com exemplo de limiar (linha vermelha) baseado nas distâncias do conjunto de validação

Numa base de dados fechada, ou seja, todas as imagens de teste pertencem a um indivíduo presente na base de dados de treinamento, o erro que pode acontecer é a classe estimada ser incorreta. Numa base de dados aberta, ou seja, uma imagem de teste pode pertencer a um indivíduo que não está presente na base de dados de treinamento, há 3 tipos de erro: uma imagem de alguém cadastrado ter a classe estimada incorretamente, uma imagem de alguém cadastrado não ter a classe estimada pois a distância mínima é maior que o limiar e uma imagem de alguém não presente na base de dados ter uma classe estimada pois a distância mínima é menor que limiar.

Capítulo 4

Experimentos e Resultados

Este capítulo tem o objetivo de descrever os experimentos realizados e resultados obtidos. Os testes foram elaborados visando demonstrar a influência da resolução nas taxas de reconhecimento facial e comparar o método proposto com algumas abordagens existentes. Para simular a situação em que a pessoa está distante da câmera e a região da sua face na imagem está em baixa resolução, as imagens usadas das bases de dados são redimensionadas via interpolação bicúbica.

O PCA é bastante usado na área de reconhecimento facial como extrator de características. Como descrito anteriormente muitas técnicas surgiram tendo como base PCA ou usando PCA numa etapa do processo. O método proposto é baseado em uma variação do PCA, o *Eigeface* introduzido por TURK e PENTLAND (1991). O objetivo dos experimentos é mostrar que os autovetores do *Eigenface*, também conhecidos como *eigenvectors*, podem ser redimensionados e mesmo assim não perdem a informação de representação do conjunto de treinamento possibilitando serem usados ainda na classificação.

Neste capítulo os experimentos comparam a Redução de Características I e II propostas com outras duas abordagens baseadas, o *Eigenface* com redução das imagens de treinamento e o *Eigenface* com aumento das imagens de teste. O *Eigenface* com redução das imagens de treinamento reduz a resolução de todas as imagens de treinamento para a mesma resolução da imagem de teste, então é aplicado o algoritmo de *Eigenface* sobre essas imagens de treinamento resultantes. O *Eigenface* com aumento das imagens de teste aplica o algoritmo de *Eigenface* sobre as imagens de treinamento com a resolução original. Para realizar o teste, as imagens de teste tem sua resolução ampliada para a resolução que foi realizado o treinamento.

Nas seções seguintes são descritas as bases de dados, protocolo de treinamento e teste e resultados.

4.1 Bases de dados

São utilizadas as seguintes bases de dados de faces durante os experimentos, ORL, Yale, Sheffield e Georgia Tech. As características de cada base de dados são descritas a seguir. Os experimentos nas bases de dados foram feitos tomando como base a resolução das imagens em 112×92 *pixels* e as reduções aplicadas com fatores de escala de $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$ e $\frac{1}{32}$, ou seja, as imagens resultantes após a redução têm resolução 56×46 , 28×23 , 14×12 , 7×6 e 4×3 .

4.1.1 ORL

A base de dados ORL contém imagens de 40 pessoas. Cada pessoa tem 10 imagens de resolução 112×92 *pixels* em 256 níveis de cinza. Esta base de dados foi montada pelo AT&T Laboratories Cambridge. As imagens contém pequenas variações de iluminação, expressão facial e alguns detalhes como óculos.



Figura 16 – Exemplo de imagens da base de dados ORL

4.1.2 Yale

A base de dados Yale contém imagens de 15 pessoas. Cada pessoa tem 11 imagens de resolução 243×320 *pixels* em 165 níveis de cinza. As imagens contém variações de expressão facial. As imagens são recortadas para 112×92 *pixels*.

4.1.3 Sheffield (UMIST)

A base de dados Sheffield contém imagens de 20 pessoas. O número de imagens por pessoa é variado, totalizando 564 imagens de resolução 220×220 *pixels* em 256 níveis de cinza. As imagens contém variações de pose de perfil até frontal. As imagens são recortadas para 112×92 *pixels*.

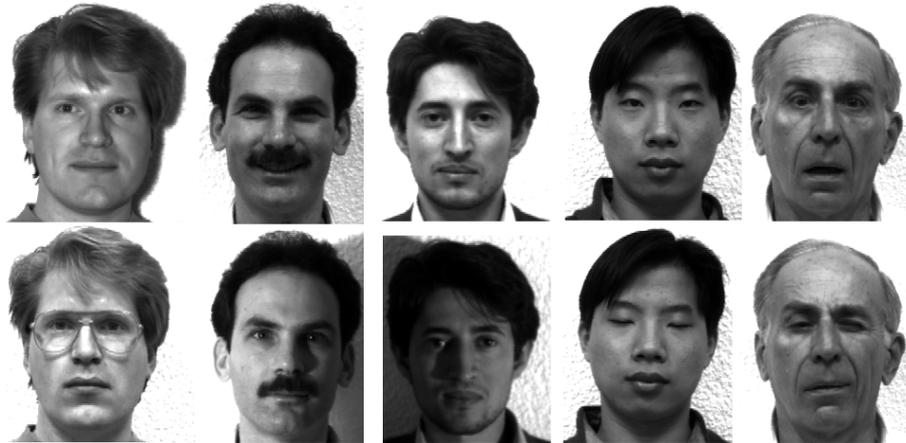


Figura 17 – Exemplo de imagens da base de dados Yale

Figura 18 – Exemplo de imagens da base de dados Sheffield



4.1.4 Georgia Tech Face Database

A base de dados Georgia Tech contém imagens de 50 pessoas com 15 imagens de cada pessoa. O tamanho médio das faces nas imagens é 150×150 pixels. As imagens estão em formato JPEG colorido e são convertidas para escala de cinza com 256 níveis. As imagens contém poses frontais e inclinadas com variações de expressões faciais, iluminação e escala. As posições das faces nas imagens são fornecidas com a base de dados.

A Tabela 1 apresenta um resumo das características das bases de dados.

4.2 Protocolo de treinamento e teste

Foram definidos aleatoriamente 100 *folds* utilizando o número de imagens da Tabela 2. Foram realizados experimentos com 4 conjuntos de treinamento, apenas usando a ORL, apenas Yale, apenas Sheffield e por fim foi realizado um experimento em que as 3 bases formaram um conjunto de treinamento. A junção das bases tem o intuito observar o

Figura 19 – Exemplo de imagens da base de dados Georgia Tech



Tabela 1 – Características presentes nas bases de dados usadas presentes no treinamento

Características	ORL	Yale	Sheffield
Iluminação	Sim	Não	Não
Expressão	Sim	Sim	Não
Óculos	Sim	Não	Sim
Pose	Não	Não	Sim

comportamento dos métodos num conjunto maior de imagens e numa grande variabilidade de situações. A base de dados Georgia Tech é usada apenas para testes com o objetivo de verificar o desempenho dos métodos quando são apresentadas imagens que não estão no conjunto de treinamento.

Tabela 2 – Número de imagens no conjunto de treinamento, validação e teste

Base de dados	Treinamento	Validação	Teste
ORL	240 (6 por pessoa)	80 (2 por pessoa)	80 (2 por pessoa)
Yale	105 (7 por pessoa)	30 (2 por pessoa)	30 (2 por pessoa)
Sheffield	220 (11 por pessoa)	80 (4 por pessoa)	80 (4 por pessoa)
Georgia Tech	0	0	750

4.3 Resultados

Há diversas métricas que relacionam os acertos e erros nos conjuntos de exemplos positivos e negativos. Exemplo positivo é uma imagem facial de um indivíduo presente na base de dados de treinamento e um exemplo negativo é uma imagem facial de uma

Tabela 3 – Representação das possibilidades de classe real e classe estimada

		Classe real	
		Positivo	Negativo
Classe estimada	Positivo	v_p	f_p
	Negativo	f_n	v_n

pessoa ausente da base de dados de treinamento. A Tabela 3 apresenta um resumo de erros e acertos (POWERS, 2007). Os acertos são chamados de verdadeiros positivos v_p e verdadeiros negativos v_n . Um verdadeiro positivo ocorre quando uma imagem facial de um indivíduo presente na base de dados de treinamento é aceita, ou seja, a distância mínima fica abaixo do limiar. Um verdadeiro negativo ocorre quando uma imagem facial de um indivíduo ausente da base de dados de treinamento é rejeitada, ou seja, a distância mínima fica acima do limiar.

Quanto aos erros, eles são chamados de falsos positivos f_p e falsos negativos f_n . Um falso positivo ocorre quando uma imagem facial de um indivíduo ausente da base de dados de treinamento é aceita. Um falso negativo ocorre quando uma imagem facial de um indivíduo presente na base de dados de treinamento é rejeitada.

Os resultados serão analisados sob métricas extraídas da Tabela 3, que são Sensibilidade, Especificidade, Precisão, Precisão Inversa e Acurácia. Por fim, os métodos são comparados quanto ao tempo de processamento.

4.3.1 Sensibilidade

A Sensibilidade dada pela Equação 4.1, que fornece a proporção de imagens de indivíduos presentes na base de dados de treinamento que são corretamente classificados.

$$\text{Sensibilidade} = \frac{v_p}{v_p + f_n} \quad (4.1)$$

Os resultados utilizando imagens da base de dados ORL são mostrados na Tabela 4 e na Figura 20. *Eigenface* com redução das imagens de treinamento e Redução de Características I têm uma queda considerável de desempenho apenas para a resolução mais baixa 4×3 . *Eigenface* com aumento das imagens de teste tem uma queda de mais de 20 pontos percentuais na média na resolução 7×6 em relação a resolução 14×12 e fica abaixo de 20% na resolução 4×3 . Redução de Características II tem uma variação de desempenho semelhante ao *Eigenface* ficando num ponto intermediário na resolução 7×6 .

Os resultados utilizando imagens da base de dados Yale são mostrados na Tabela 5 e na Figura 21. A Sensibilidade tem uma queda de desempenho mais suave nas resoluções mais baixas para *Eigenface* com redução das imagens de treinamento e Redução de Características I. *Eigenface* com aumento das imagens de teste possui uma queda de

Tabela 4 – Sensibilidade por resolução com imagens de treinamento, validação e teste da base de dados ORL. Média μ e desvio padrão σ em %.

Resolução	<i>Eigenface</i> com redução das imagens de treinamento	Redução de características I	<i>Eigenface</i> com aumento das imagens de teste	Redução de características II
	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
4×3	30.6 ± 9.0	33.9 ± 8.1	16.6 ± 5.9	13.7 ± 4.4
7×6	87.2 ± 4.4	87.2 ± 4.6	67.6 ± 6.8	79.6 ± 4.2
14×12	88.8 ± 4.1	89.2 ± 4.1	89.7 ± 3.6	89.4 ± 4.1
28×23	91.2 ± 3.1	91.4 ± 2.9	92.0 ± 3.5	91.5 ± 3.4
56×46	91.5 ± 3.1	91.3 ± 3.3	91.3 ± 3.7	91.4 ± 3.8
112×92	91.1 ± 3.8	91.1 ± 3.8	91.1 ± 3.8	91.1 ± 3.8

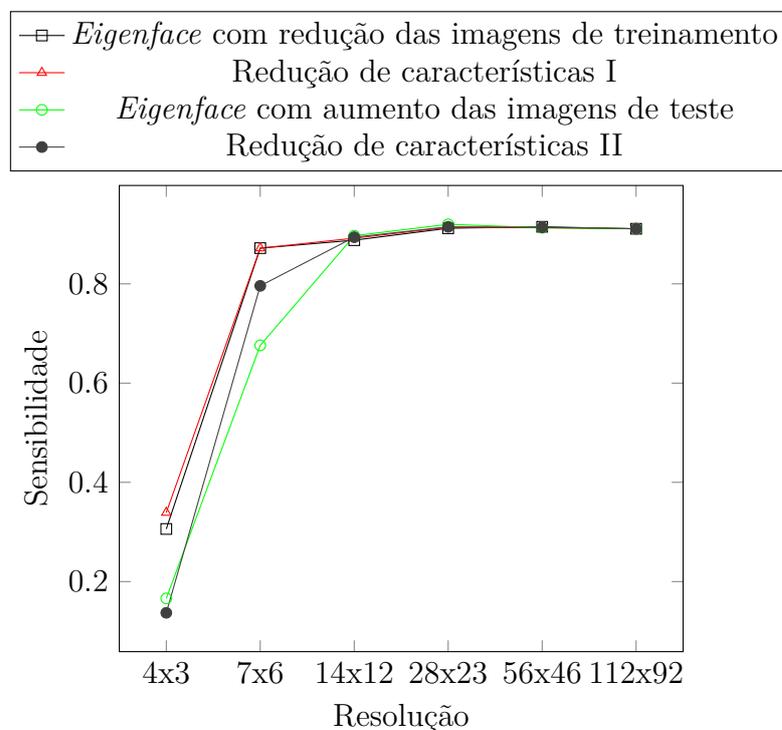


Figura 20 – Sensibilidade por resolução com imagens de treinamento, validação e teste da base de dados ORL

desempenho bem acentuada acertando apenas 7,7% das imagens enquanto na Redução de Características I o acerto é de 58,3% para a resolução 4×3 . Redução de Características II tem uma queda de desempenho em relação as outras reduções apenas na resolução 4×3 .

Os resultados utilizando imagens da base de dados Sheffield são mostrados na Tabela 6 e na Figura 22. Com características semelhantes ao resultado na base de dados Yale, a Sensibilidade para *Eigenface* com redução das imagens de treinamento. Redução

Tabela 5 – Sensibilidade por resolução com imagens de treinamento, validação e teste da base de dados Yale. Média μ e desvio padrão σ em %.

Resolução	<i>Eigenface</i> com redução das imagens de treinamento	Redução de características I	<i>Eigenface</i> com aumento das imagens de teste	Redução de características II
	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
4×3	56.6 ± 10.5	58.3 ± 9.9	7.7 ± 5.1	29.8 ± 8.2
7×6	66.2 ± 9.6	67.2 ± 8.8	47.7 ± 12.6	66.3 ± 8.2
14×12	66.3 ± 9.3	67.5 ± 8.9	66.7 ± 9.1	69.3 ± 8.2
28×23	70.6 ± 8.7	71.0 ± 8.6	71.6 ± 8.3	72.4 ± 8.4
56×46	72.1 ± 8.2	71.9 ± 8.4	71.6 ± 8.4	71.9 ± 8.5
112×92	71.4 ± 8.4	71.4 ± 8.4	71.4 ± 8.4	71.4 ± 8.4

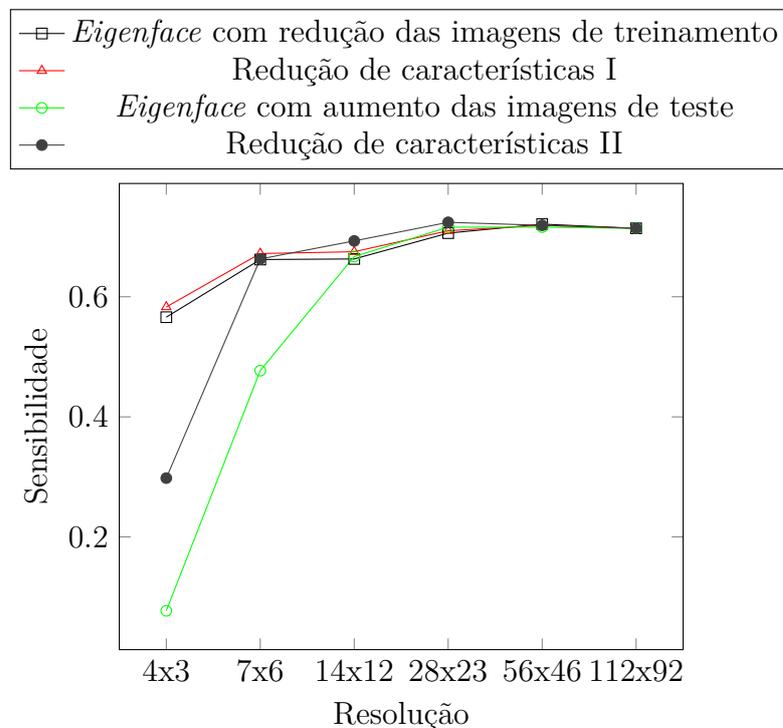


Figura 21 – Sensibilidade por resolução com imagens de treinamento, validação e teste da base de dados Yale

de Características I e II têm uma queda desempenho apenas para a resolução mais baixa 4×3 . *Eigenface* com aumento das imagens de teste tem uma queda de mais de 20 pontos percentuais na média na resolução 7×6 em relação a resolução 14×12 .

Os resultados juntando as 3 bases de dados para um conjunto de treinamento são mostrados na Tabela 7 e na Figura 23. A Sensibilidade para *Eigenface* com redução das imagens de treinamento e Redução de Características I apresentam desempenhos semelhantes, enquanto *Eigenface* com aumento das imagens de teste fica com desempenho

Tabela 6 – Sensibilidade por resolução com imagens de treinamento, validação e teste da base de dados Sheffield. Média μ e desvio padrão σ em %.

Resolução	<i>Eigenface</i> com redução das imagens de treinamento	Redução de características I	<i>Eigenface</i> com aumento das imagens de teste	Redução de características II
	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
4×3	49.0 ± 11.3	52.7 ± 13.3	18.7 ± 4.6	34.9 ± 6.2
7×6	88.0 ± 4.7	88.2 ± 4.9	63.1 ± 8.1	86.4 ± 4.4
14×12	87.6 ± 4.6	87.3 ± 4.8	87.2 ± 4.6	86.4 ± 4.8
28×23	90.8 ± 3.5	90.7 ± 3.6	91.2 ± 3.4	90.8 ± 3.6
56×46	90.8 ± 3.5	90.8 ± 3.5	90.9 ± 3.5	90.8 ± 3.5
112×92	90.9 ± 3.5	90.9 ± 3.5	90.9 ± 3.5	90.9 ± 3.5

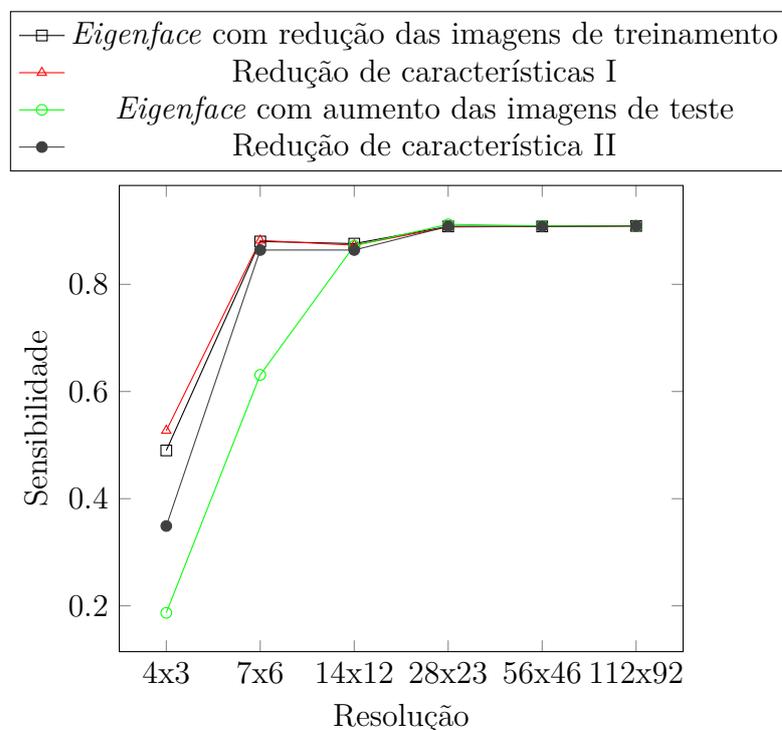


Figura 22 – Sensibilidade por resolução com imagens de treinamento, validação e teste da base de dados Sheffield

inferior nas resoluções mais baixas. Redução de Características II tem um desempenho intermediário para resolução 7×6 e fica abaixo de todos na resolução 4×3 .

4.3.2 Especificidade

A Especificidade é dada pela Equação 4.2, que fornece a proporção de imagens de indivíduos que não estão presentes na base de dados de treinamento e são corretamente

Tabela 7 – Sensibilidade por resolução com imagens de treinamento das bases de dados ORL, Yale e Sheffield. Média μ e desvio padrão σ em %.

Resolução	<i>Eigenface</i> com redução das imagens de treinamento	Redução de características I	<i>Eigenface</i> com aumento das imagens de teste	Redução de características II
	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
4×3	49.7 ± 6.6	56.8 ± 6.9	21.6 ± 3.0	12.9 ± 2.9
7×6	86.1 ± 2.7	86.9 ± 2.6	69.5 ± 3.7	79.5 ± 2.5
14×12	86.1 ± 2.5	86.5 ± 2.5	85.7 ± 2.3	86.0 ± 2.6
28×23	88.5 ± 2.1	88.5 ± 1.9	88.4 ± 2.2	88.5 ± 2.1
56×46	88.5 ± 2.0	88.4 ± 2.1	88.4 ± 2.2	88.4 ± 2.1
112×92	88.4 ± 2.1	88.4 ± 2.1	88.4 ± 2.1	88.4 ± 2.1

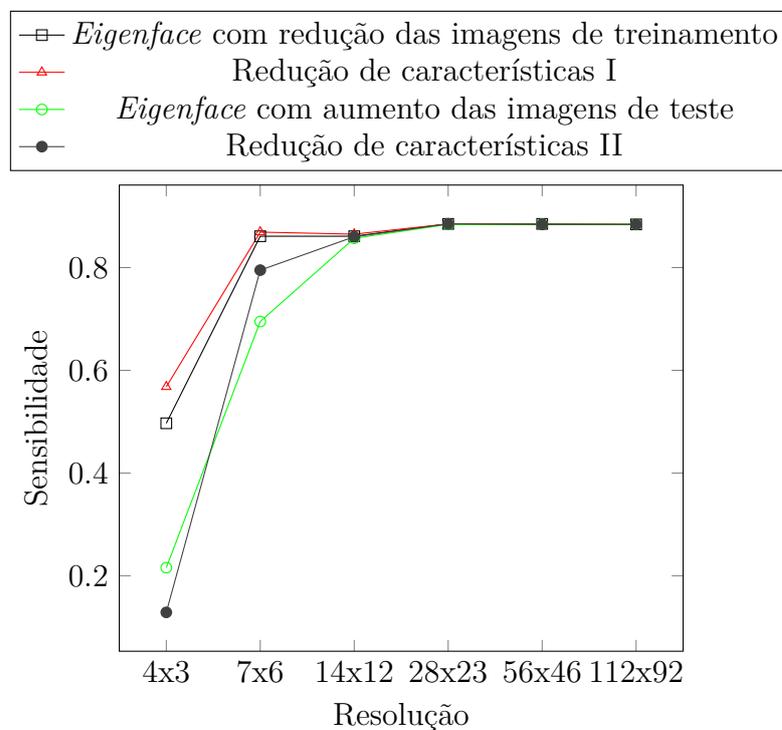


Figura 23 – Sensibilidade por resolução com imagens de treinamento das bases de dados ORL, Yale e Sheffield

não classificados, ou seja, não lhe atribuída classe de um indivíduo treinado.

$$\text{Especificidade} = \frac{v_n}{f_p + v_n} \quad (4.2)$$

Foram realizados experimentos usando imagens faciais de pessoas não presentes o treinamento. Para avaliar o desempenho foi utilizado o algoritmo descrito na Seção 3.3. Se a saída do método for um ID de algum indivíduo presente no treinamento a resposta é um erro chamado de falso positivo, caso a saída seja que a face é desconhecida a resposta é

um acerto chamado de verdadeiro negativo.

Na Tabela 8 e Figura 24 são mostrados os resultados para o treinamento realizado com imagens da base de dados ORL.

Tabela 8 – Especificidade por resolução com imagens de treinamento e validação da base de dados ORL e imagens de teste da base de dados Georgia Tech. Média μ e desvio padrão σ em %.

Resolução	<i>Eigenface</i> com redução das imagens de treinamento	Redução de características I	<i>Eigenface</i> com aumento das imagens de teste	Redução de características II
	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
4×3	92.6 ± 3.6	93.6 ± 2.8	55.8 ± 17.5	42.3 ± 15.1
7×6	93.6 ± 4.8	94.9 ± 4.1	62.2 ± 15.3	55.6 ± 7.7
14×12	93.2 ± 5.0	93.8 ± 4.4	74.8 ± 11.9	82.5 ± 7.6
28×23	93.1 ± 5.5	92.8 ± 5.7	82.5 ± 9.7	88.0 ± 8.2
56×46	89.8 ± 6.8	90.9 ± 5.6	84.7 ± 9.0	86.1 ± 8.8
112×92	85.2 ± 8.7	85.2 ± 8.7	85.2 ± 8.7	85.2 ± 8.7

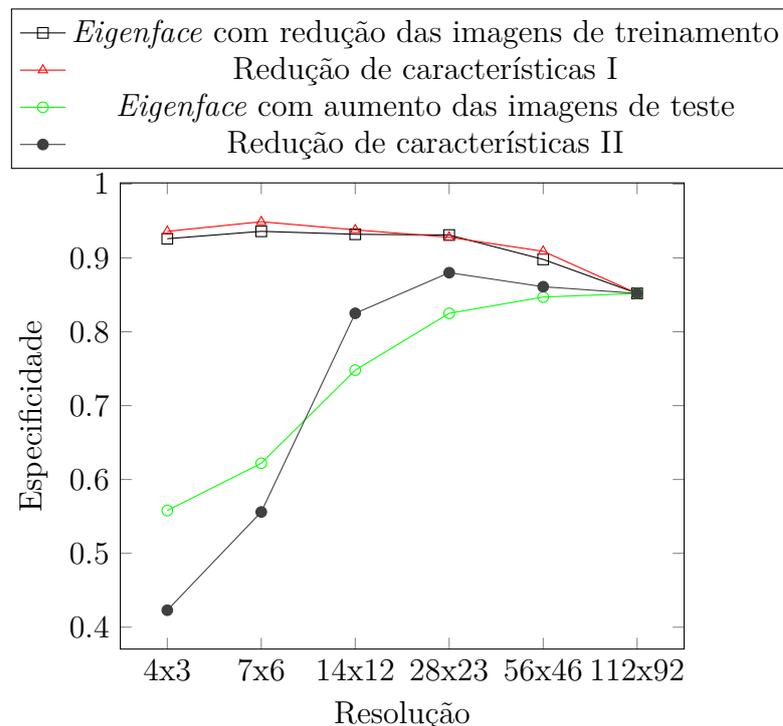


Figura 24 – Especificidade por resolução com imagens de treinamento e validação da base de dados ORL e imagens de teste da base de dados Georgia Tech

Eigenface com redução das imagens de treinamento e Redução de características I apresentam uma maior taxa de acerto nas baixas resoluções. Para *Eigenface* com aumento das imagens de teste acontece o contrário errando mais nas baixas resoluções. Redução

de Características II tem um desempenho intermediário nas resoluções entre 14×12 e 56×46 e um pior desempenho nas resoluções 4×3 e 7×6 .

Na Tabela 9 e Figura 25 são mostrados os resultados para o treinamento realizado com imagens da base de dados Yale.

Tabela 9 – Especificidade por resolução com imagens de treinamento e validação da base de dados Yale e imagens de teste da base de dados Georgia Tech. Média μ e desvio padrão σ em %.

Resolução	<i>Eigenface</i> com redução das imagens de treinamento	Redução de características I	<i>Eigenface</i> com aumento das imagens de teste	Redução de características II
	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
4×3	94.2 ± 14.2	94.4 ± 12.3	60.6 ± 20.2	24.7 ± 24.2
7×6	96.7 ± 8.5	96.8 ± 7.3	74.5 ± 19.8	78.3 ± 17.5
14×12	90.8 ± 13.5	90.3 ± 12.5	78.4 ± 19.1	80.5 ± 19.3
28×23	86.1 ± 15.4	84.5 ± 16.3	79.4 ± 18.8	80.5 ± 18.8
56×46	80.3 ± 18.3	79.5 ± 18.6	79.5 ± 18.7	78.5 ± 19.3
112×92	79.5 ± 18.7	79.5 ± 18.7	79.5 ± 18.7	79.5 ± 18.7

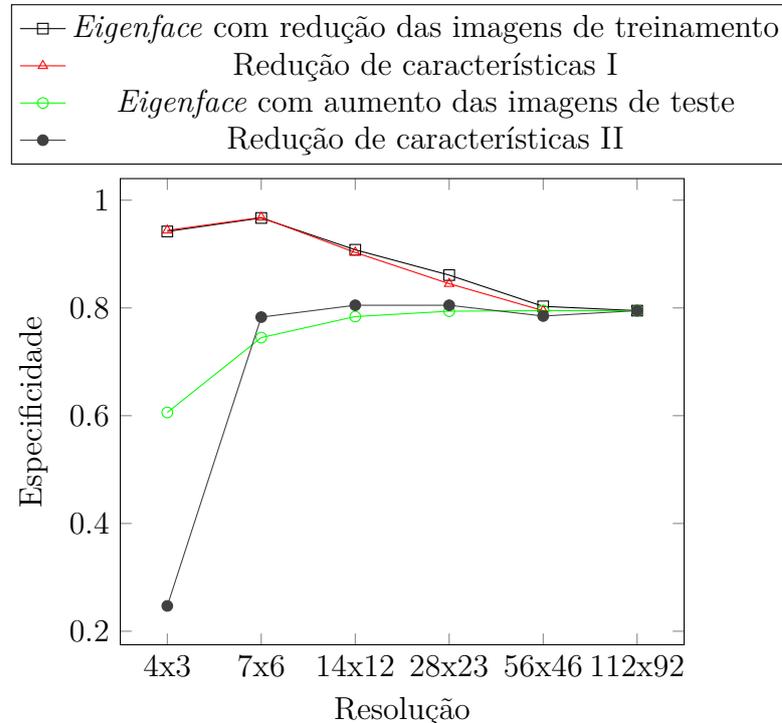


Figura 25 – Especificidade por resolução com imagens de treinamento e validação da base de dados Yale e imagens de teste da base de dados Georgia Tech

Semelhante ao resultado anterior, *Eigenface* com redução das imagens de treinamento e Redução de características I apresentam uma maior taxa de acerto nas baixas

resoluções. Para *Eigenface* com aumento das imagens de teste acontece o contrário errando mais nas baixas resoluções. Redução de Características II tem um desempenho semelhante ao *Eigenface*, tendo uma queda de desempenho significativa na resolução 4×3 .

Na Tabela 10 e Figura 26 são mostrados os resultados para o treinamento realizado com imagens da base de dados Sheffield.

Tabela 10 – Especificidade por resolução com imagens de treinamento e validação da base de dados Sheffield e imagens de teste da base de dados Georgia Tech. Média μ e desvio padrão σ em %.

Resolução	<i>Eigenface</i> com redução das imagens de treina- mento	Redução de característi- cas I	<i>Eigenface</i> com aumento das imagens de teste	Redução de característi- cas II
	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
4×3	91.8 ± 6.1	96.3 ± 3.5	72.8 ± 17.9	26.8 ± 16.5
7×6	92.7 ± 9.3	97.0 ± 6.0	73.1 ± 17.8	68.9 ± 16.3
14×12	93.4 ± 9.2	95.2 ± 9.4	81.8 ± 15.4	94.7 ± 7.4
28×23	89.2 ± 11.1	90.4 ± 11.2	83.1 ± 14.4	88.5 ± 12.2
56×46	85.7 ± 13.2	85.9 ± 13.3	83.9 ± 14.0	84.8 ± 13.8
112×92	84.0 ± 14.0	84.0 ± 14.0	84.0 ± 14.0	84.0 ± 14.0

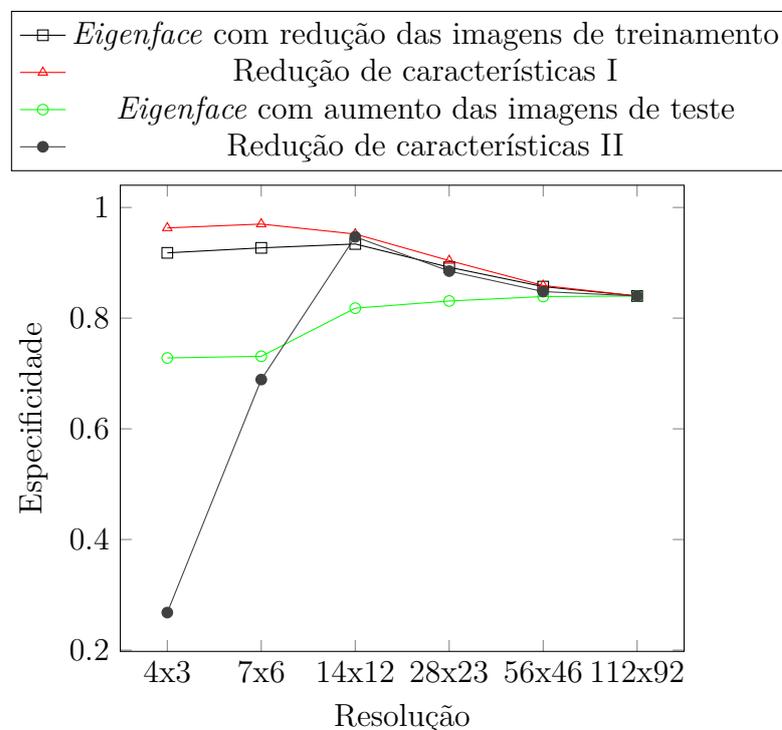


Figura 26 – Especificidade por resolução por resolução com imagens de treinamento e validação da base de dados Sheffield e imagens de teste da base de dados Georgia Tech

Novamente *Eigenface* com aumento de teste erra mais nas baixas resoluções. Nessa base de dados, Redução das características I tem um melhor desempenho que *Eigenface* com redução das imagens de treinamento. Redução de Características II tem um resultado semelhante a Redução de características I entre as resoluções 14×12 e 56×46 e o desempenho cai para as resoluções 4×3 e 7×6 .

Na Tabela 11 e Figura 27 são mostrados os resultados para o treinamento realizado com imagens das bases de dados ORL, Yale e Sheffield.

Tabela 11 – Especificidade por resolução com imagens de treinamento das bases de dados ORL, Yale e Sheffield Média μ e desvio padrão σ em %.

Resolução	<i>Eigenface</i> com redução das imagens de treina- mento	Redução de característi- cas I	<i>Eigenface</i> com aumento das imagens de teste	Redução de característi- cas II
	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
4×3	71.6 ± 9.4	72.9 ± 10.7	9.1 ± 8.2	0.5 ± 1.4
7×6	64.2 ± 18.0	60.7 ± 16.9	11.2 ± 10.6	1.0 ± 0.8
14×12	54.7 ± 18.1	54.4 ± 18.7	24.0 ± 17.2	23.6 ± 13.4
28×23	44.3 ± 21.1	42.1 ± 20.3	30.9 ± 20.8	32.1 ± 20.9
56×46	36.9 ± 22.3	35.8 ± 21.5	33.9 ± 21.8	34.1 ± 21.3
112×92	34.3 ± 21.9	34.3 ± 21.9	34.3 ± 21.9	34.3 ± 21.9

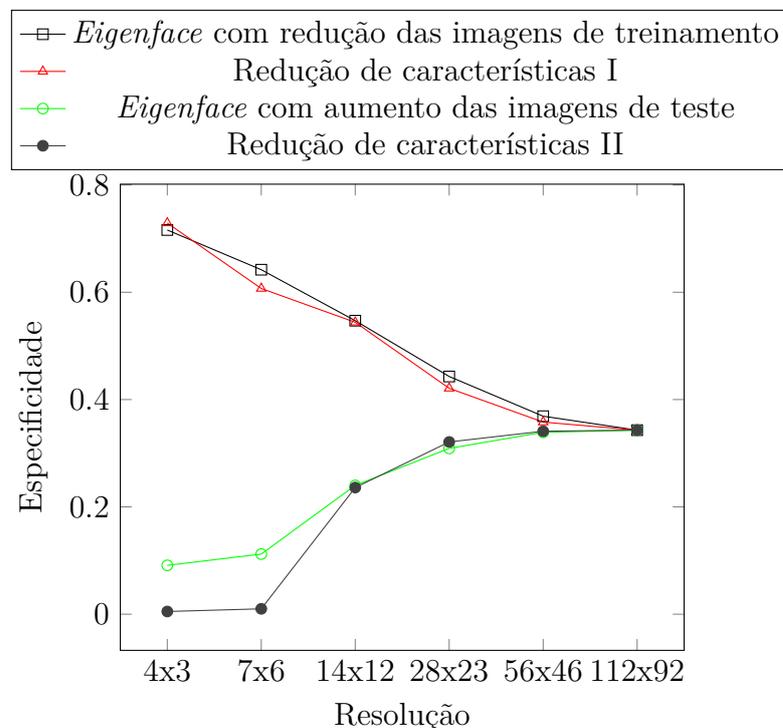


Figura 27 – Especificidade por resolução com imagens de treinamento das bases de dados ORL, Yale e Sheffield e imagens de teste da base de dados Georgia Tech

A variação das taxas pela resolução é semelhante aos resultados anteriores. *Eigenface* com aumento das imagens de teste chega a ter uma taxa de verdadeiro negativo média menor que 10%, enquanto a Redução de Características chega a média superior a 72%. O desempenho de Redução de Características II foi semelhante ao *Eigenface* com aumento das imagens de teste.

Pode-se notar que em resoluções menores é maior o acerto ao rejeitar faces de indivíduos que não estão na base de dados de treinamento para os métodos *Eigenface* com redução das imagens de treinamento, Redução de características I. Uma explicação é que nesses métodos torna-se mais difícil que uma face seja aceita. Isso pode ser na diminuição das taxas de aceitação de indivíduos presentes na base de dados, analisadas anteriormente. Os dois métodos citados então são os que se adaptam melhor ao ajuste de limiar proposto, que visa minimizar a queda de acerto de verdadeiros e ao mesmo tempo maximizar o número de verdadeiros negativos.

4.3.3 Precisão

A medida de Precisão é dada pela Equação 4.3, que fornece a proporção de exemplos classificados como positivos e estão classificados corretamente.

$$Precisao = \frac{v_p}{v_p + f_p} \quad (4.3)$$

Os resultados de Precisão são mostrados nas Figuras 28, 29, 30 e 31. Redução de Características I nessa métrica tem uma maior superioridade na base de dados Sheffield.

4.3.4 Precisão Inversa

A medida de Precisão Inversa é dada pela Equação 4.4, que fornece a proporção de exemplos classificados como negativos e estão classificados corretamente.

$$PrecisaoInversa = \frac{v_n}{v_n + f_n} \quad (4.4)$$

Os resultados de Precisão Inversa são mostrados nas Figuras 32, 33, 34 e 35. Nessa métrica, Redução de Características I tem um desempenho semelhante a *Eigenface* com redução das imagens de treinamento em todas as bases de dados.

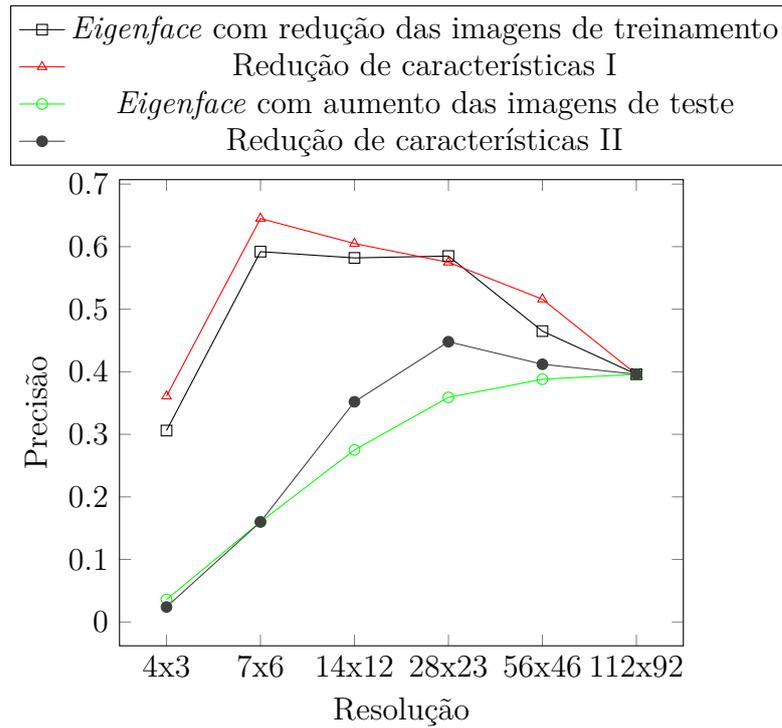


Figura 28 – Precisão por resolução com imagens de treinamento e validação da base de dados ORL e imagens de teste das bases de dados ORL e Georgia Tech

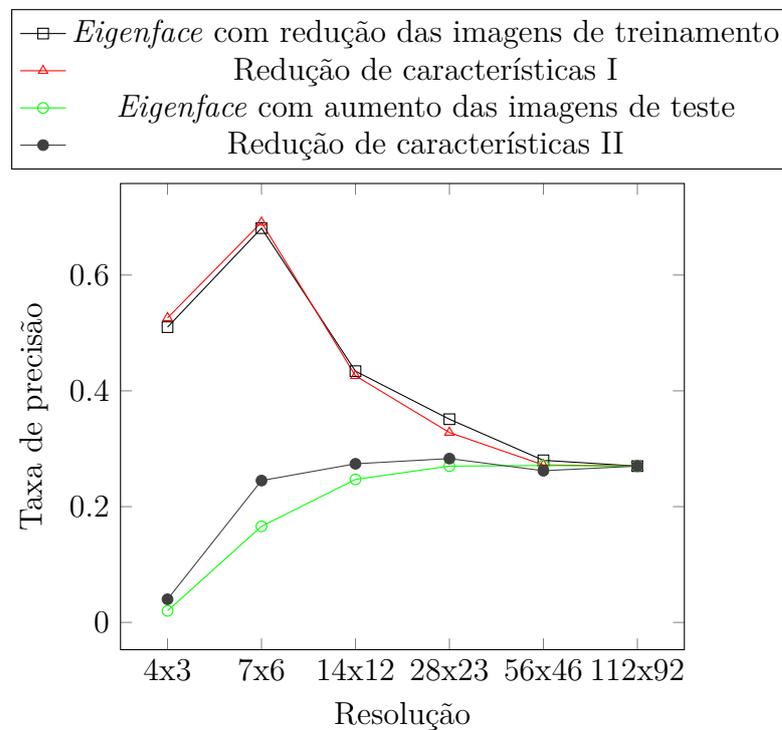


Figura 29 – Taxa de precisão por resolução com imagens de treinamento e validação da base de dados Yale e imagens de teste das bases de dados Yale e Georgia Tech

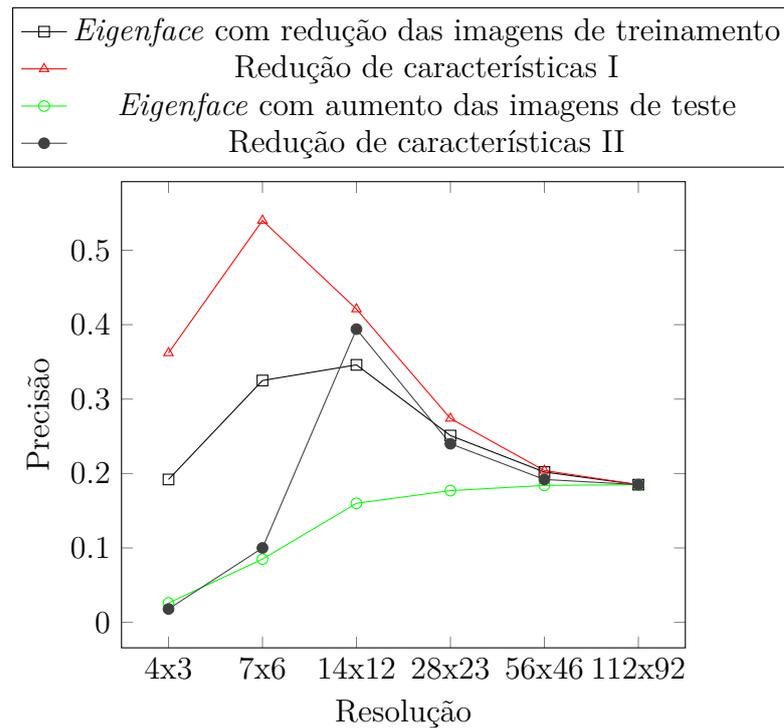


Figura 30 – Precisão por resolução com imagens de treinamento e validação da base de dados Sheffield e imagens de teste das bases de dados Sheffield e Georgia Tech

4.3.5 Acurácia

A medida de Acurácia é dada pela Equação 4.5, que fornece a proporção de exemplos classificados corretamente.

$$Acuracia = \frac{v_p + v_n}{v_p + f_p + v_n + f_n} \quad (4.5)$$

Os resultados de acurácia são mostrados nas Figuras 36, 37, 38 e 39. Para a Acurácia, o desempenho de Redução de Características I tem maior superioridade na base de dados Sheffield.

4.3.6 Tempo de processamento

Por fim os métodos são analisados quanto ao tempo de processamento das imagens. Os resultados são mostrados na Tabelas 12, 13, 14 e 15.

Comparando os tempos médios de processamento, o método com menor média de tempo de processamento é *Eigenface* com redução das imagens de treinamento. Em seguida, Redução de Características I, Redução de Características II e *Eigenface* com aumento das imagens de teste, respectivamente. Pode-se dizer que há uma evolução quanto ao tempo de processamento em comparação com *Eigenface* com aumento das imagens

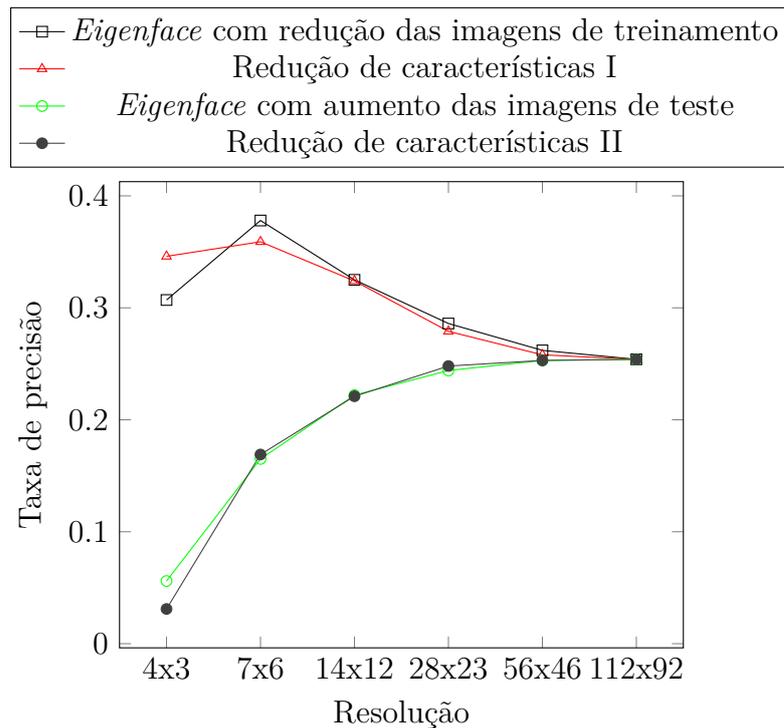


Figura 31 – Precisão por resolução com imagens de treinamento e validação das bases de dados ORL, Yale e Sheffield e imagens de teste das bases de dados ORL, Yale, Sheffield e Georgia Tech

Tabela 12 – Tempo médio de processamento do treinamento, validação e teste da base de dados ORL. Média μ e desvio padrão σ em segundos.

Resolução	<i>Eigenface</i> com redução das imagens de treinamento	<i>Eigenface</i> com aumento das imagens de teste	Redução de Características I	Redução de Características II
	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
4 × 3	1.18 ± 0.05	4.14 ± 0.20	2.47 ± 0.03	2.67 ± 0.10
7 × 6	1.31 ± 0.07	4.00 ± 0.15	2.55 ± 0.09	2.65 ± 0.14
14 × 12	1.24 ± 0.06	4.02 ± 0.63	2.49 ± 0.20	2.52 ± 0.25
28 × 23	1.48 ± 0.07	4.14 ± 0.13	2.71 ± 0.14	2.85 ± 0.13
56 × 46	1.94 ± 0.09	4.17 ± 0.21	3.29 ± 0.14	3.46 ± 0.11
112 × 92	1.83 ± 0.08	2.07 ± 0.14	5.13 ± 0.22	5.30 ± 0.22

de teste. Uma justificativa para a superioridade do método *Eigenface* com redução das imagens de treinamento, se deve por realizar o treinamento com imagens pequenas, que consequente requer menos processamento e memória nos cálculos da matriz de covariância, autovalores e autovetores.

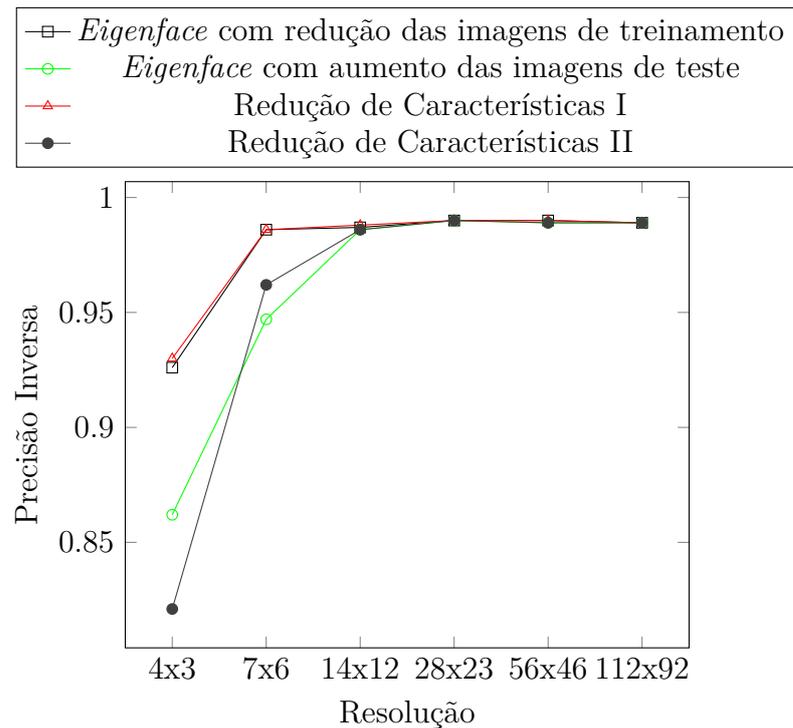


Figura 32 – Precisão Inversa por resolução com imagens de treinamento e validação da base de dados ORL e imagens de teste das bases de dados ORL e Georgia Tech

Tabela 13 – Tempo médio de processamento do treinamento, validação e teste da base de dados Yale. Média μ e desvio padrão σ em segundos.

Resolução	<i>Eigenface</i> com redução das imagens de treinamento	<i>Eigenface</i> com aumento das imagens de teste	Redução de Características I	Redução de Características II
	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
4 × 3	0.51 ± 0.08	2.95 ± 0.14	1.45 ± 0.21	1.55 ± 0.15
7 × 6	0.55 ± 0.02	2.90 ± 0.12	1.53 ± 0.05	1.57 ± 0.07
14 × 12	0.52 ± 0.01	2.83 ± 0.10	1.46 ± 0.04	1.52 ± 0.08
28 × 23	0.60 ± 0.03	2.94 ± 0.12	1.68 ± 0.08	1.74 ± 0.16
56 × 46	0.69 ± 0.03	3.02 ± 0.09	2.03 ± 0.12	2.04 ± 0.06
112 × 92	0.54 ± 0.05	1.12 ± 0.08	2.92 ± 0.13	3.00 ± 0.17

4.4 Experimentos adicionais

4.4.1 Interpolação bilinear

Um método alternativo para redimensionamento de imagens é a interpolação bilinear (GONZALEZ e WOODS, 2000). Nesta Seção os resultados de Redução de Características I e II são comparados se usados com interpolação bicúbica e com interpolação bilinear. Os

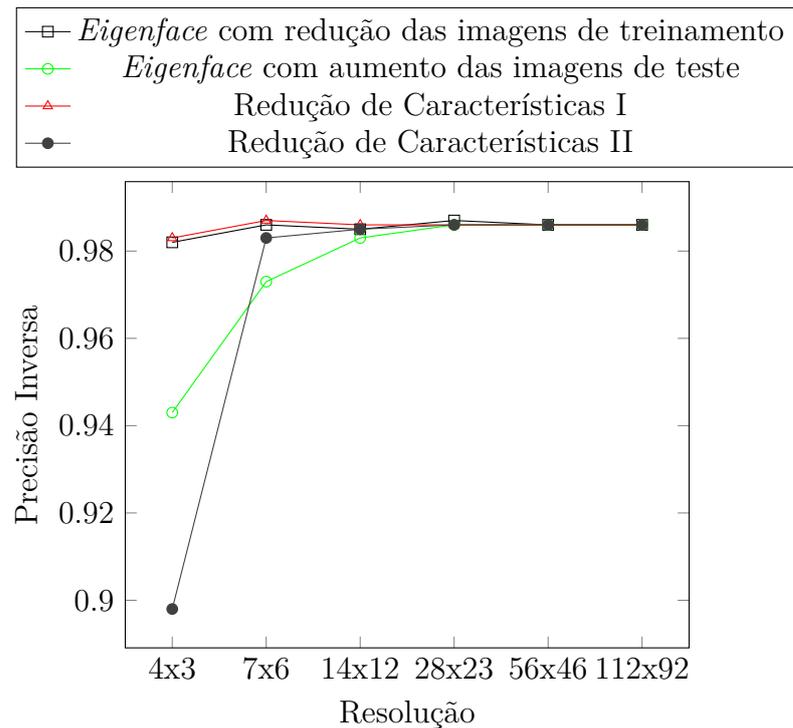


Figura 33 – Precisão Inversa por resolução com imagens de treinamento e validação da base de dados Yale e imagens de teste das bases de dados Yale e Georgia Tech

Tabela 14 – Tempo médio de processamento do treinamento, validação e teste da base de dados Sheffield. Média μ e desvio padrão σ em segundos.

Resolução	<i>Eigenface</i> com redução das imagens de treinamento	<i>Eigenface</i> com aumento das imagens de teste	Redução de Características I	Redução de Características II
	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
4 × 3	1.05 ± 0.04	3.36 ± 0.15	1.92 ± 0.07	2.06 ± 0.08
7 × 6	1.18 ± 0.03	3.40 ± 0.11	2.01 ± 0.06	2.10 ± 0.07
14 × 12	1.08 ± 0.06	3.47 ± 0.14	1.97 ± 0.10	2.08 ± 0.14
28 × 23	1.30 ± 0.06	3.45 ± 0.11	2.11 ± 0.10	2.24 ± 0.08
56 × 46	1.49 ± 0.04	3.50 ± 0.16	2.56 ± 0.10	2.72 ± 0.13
112 × 92	1.19 ± 0.08	1.59 ± 0.07	3.82 ± 0.12	3.93 ± 0.19

resultados são mostrados nas Figuras 40, 41, 42, 43 e 44.

Pode-se perceber que as duas interpolações tem resultados similares no conjunto de exemplos positivos, analisados pela Sensibilidade. Já no conjunto de exemplos negativos a interpolação bilinear tem um desempenho superior, analisado pela Especificidade. Esse tipo de interpolação pode ser usado em trabalhos futuros, já mostrou um potencial de melhora na rejeição de pessoas não cadastradas.

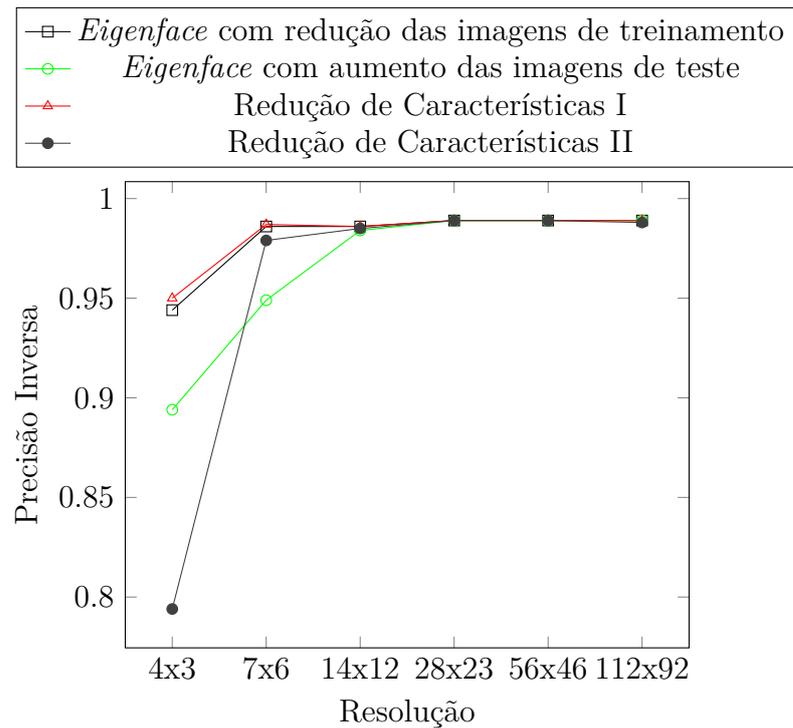


Figura 34 – Precisão Inversa por resolução com imagens de treinamento e validação da base de dados Sheffield e imagens de teste das bases de dados Sheffield e Georgia Tech

Tabela 15 – Tempo médio de processamento do treinamento, validação e teste da base de dados ORL, Yale e Sheffield. Média μ e desvio padrão σ em segundos.

Resolução	<i>Eigenface</i> com redução das imagens de treinamento	<i>Eigenface</i> com aumento das imagens de teste	Redução de Características I	Redução de Características II
	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
4 × 3	2.72 ± 0.12	5.36 ± 0.25	3.59 ± 0.18	3.93 ± 0.26
7 × 6	2.91 ± 0.10	5.22 ± 0.21	3.48 ± 0.28	3.87 ± 0.17
14 × 12	2.75 ± 0.10	5.24 ± 0.26	3.53 ± 0.18	3.86 ± 0.10
28 × 23	3.95 ± 0.17	5.29 ± 0.24	3.73 ± 0.16	4.14 ± 0.20
56 × 46	4.88 ± 0.19	5.37 ± 0.20	4.42 ± 0.20	4.79 ± 0.24
112 × 92	4.69 ± 0.20	3.26 ± 0.20	6.15 ± 0.28	6.15 ± 0.15

4.5 Considerações

Em resumo, nos experimentos de identificação, que as imagens de teste pertencem a algum indivíduo presente no treinamento, Redução das Características I teve um desempenho semelhante ou superior a *Eigenface* com Redução das imagens de treinamento e superior a *Eigenface* com aumento das imagens de teste. Redução de Características II

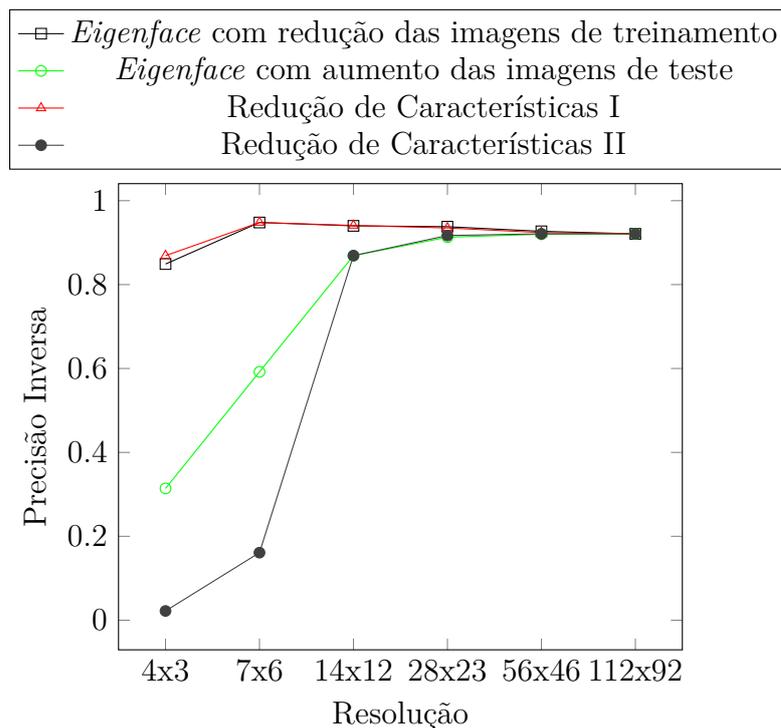


Figura 35 – Precisão Inversa por resolução com imagens de treinamento e validação das bases de dados ORL, Yale e Sheffield e imagens de teste das bases de dados ORL, Yale, Sheffield e Georgia Tech

teve um desempenho semelhante a Redução de Características I nas resoluções médias tendo uma piora desempenho na resolução mais baixa testada.

Nos experimentos que a imagem de teste é de alguém não presente no treinamento, Redução de característica I foi também superior em todos os casos a *Eigenface* com aumento das imagens de teste. Redução de Características I foi superior a *Eigenface* com redução das imagens de treinamento nas bases de dados ORL e Sheffield e um pouco inferior na base de dados Yale e na junção das 3 bases de dados. Redução de Características II teve um desempenho intermediário entre Redução de Características I e *Eigenface* com aumento das imagens de teste na maioria das bases para resoluções médias e um desempenho fraco na menor resolução testada.

Analisando os gráficos de Precisão, vemos que o desempenho de Redução de Características I teve uma maior diferença em relação a *Eigenface* com redução das imagens de treinamento na base de dados Sheffield. *Eigenface* com aumento das imagens de teste teve uma queda de desempenho para maiores resoluções em todas as bases de dados. Nos pontos em que o desempenho de Redução de Características I não foi superior ficou bem próximo a *Eigenface* com redução das imagens de treinamento. Redução de Características II teve um desempenho próximo a *Eigenface* com aumento das imagens de teste e em algumas resoluções o superando.

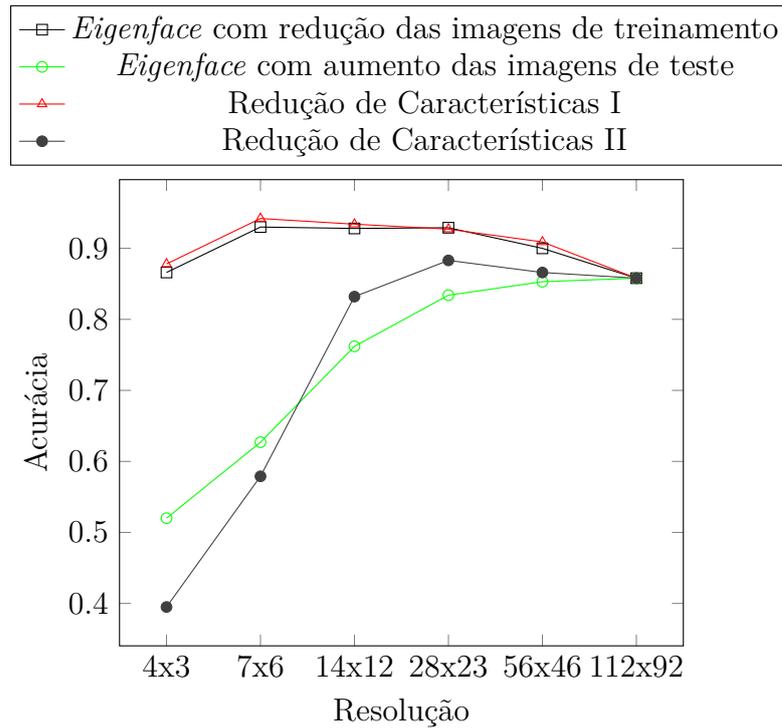


Figura 36 – Acurácia por resolução com imagens de treinamento e validação da base de dados ORL e imagens de teste das bases de dados ORL e Georgia Tech

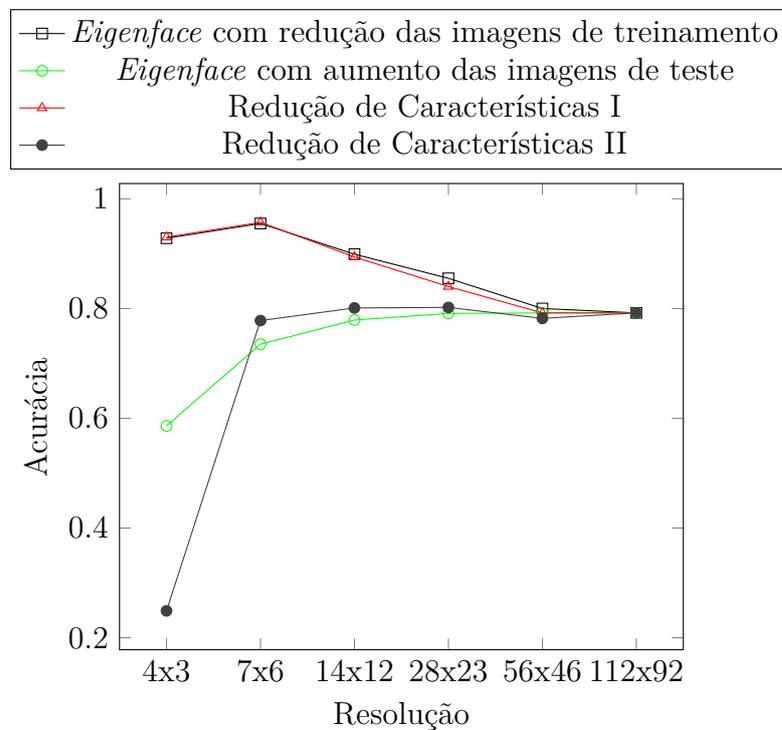


Figura 37 – Acurácia por resolução com imagens de treinamento e validação da base de dados Yale e imagens de teste das bases de dados Yale e Georgia Tech

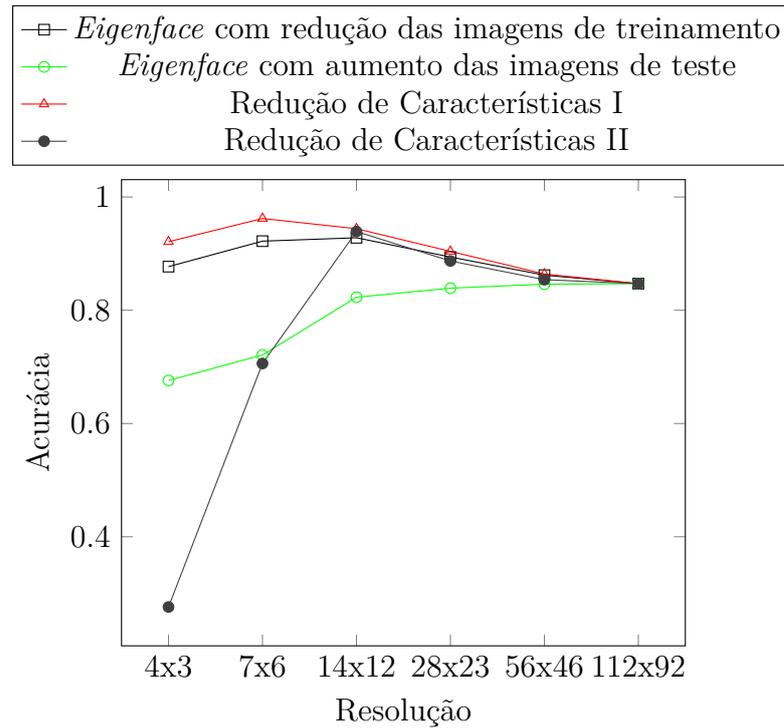


Figura 38 – Acurácia por resolução com imagens de treinamento e validação da base de dados Sheffield e imagens de teste das bases de dados Sheffield e Georgia Tech

Para a Precisão Inversa, Redução de Característica I e *Eigenface* com redução das imagens de treinamento tiveram resultados muito semelhantes. Para a Acurácia, o desempenho de Redução de Característica I foi superior a *Eigenface* com redução das imagens de treinamento para as base de dados ORL e Sheffield. Redução de Características II teve desempenho semelhante a *Eigenface* com aumento das imagens de teste em resoluções maiores, superou na maioria das vezes na resolução 7×6 .

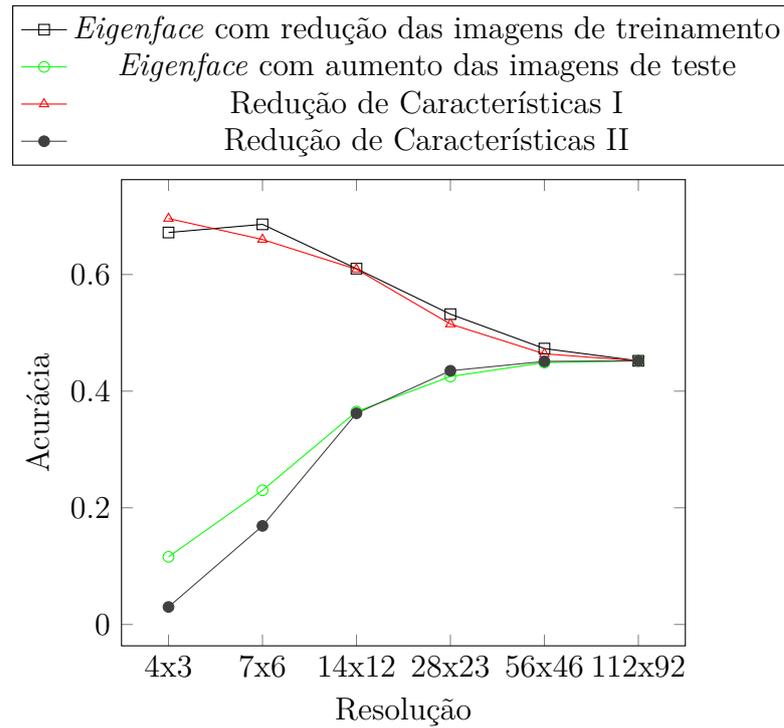


Figura 39 – Acurácia por resolução com imagens de treinamento e validação das bases de dados ORL, Yale e Sheffield e imagens de teste das bases de dados ORL, Yale, Sheffield e Georgia Tech

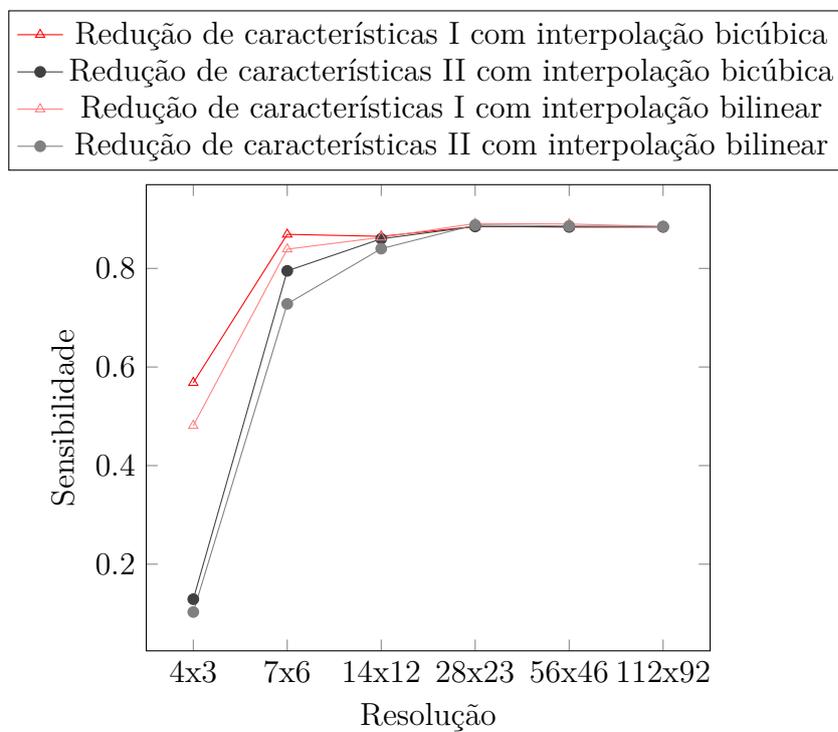


Figura 40 – Sensibilidade por resolução com imagens de treinamento das bases de dados ORL, Yale e Sheffield

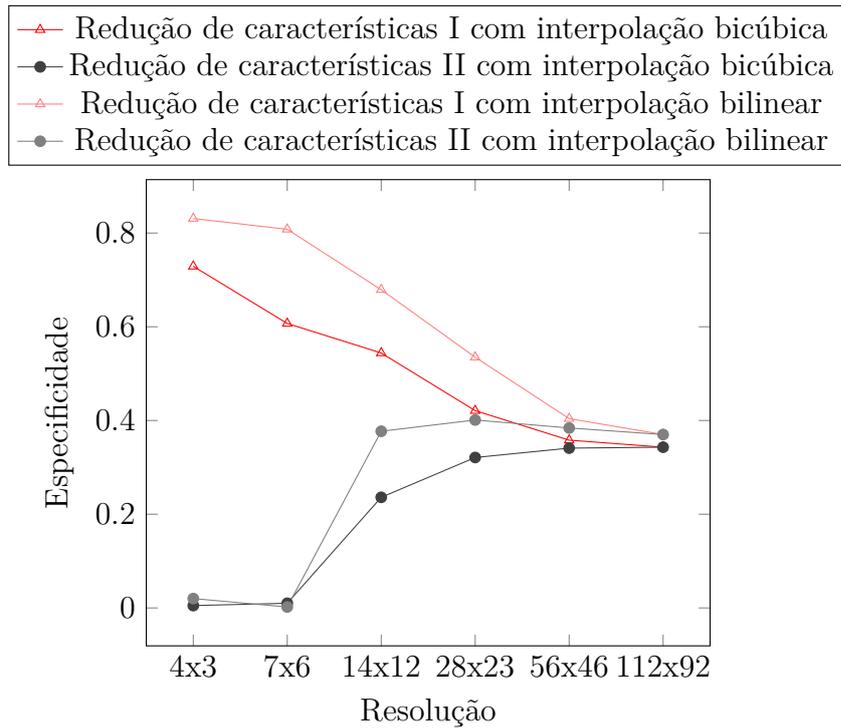


Figura 41 – Especificidade por resolução com imagens de treinamento das bases de dados ORL, Yale e Sheffield e imagens de teste da base de dados Georgia Tech

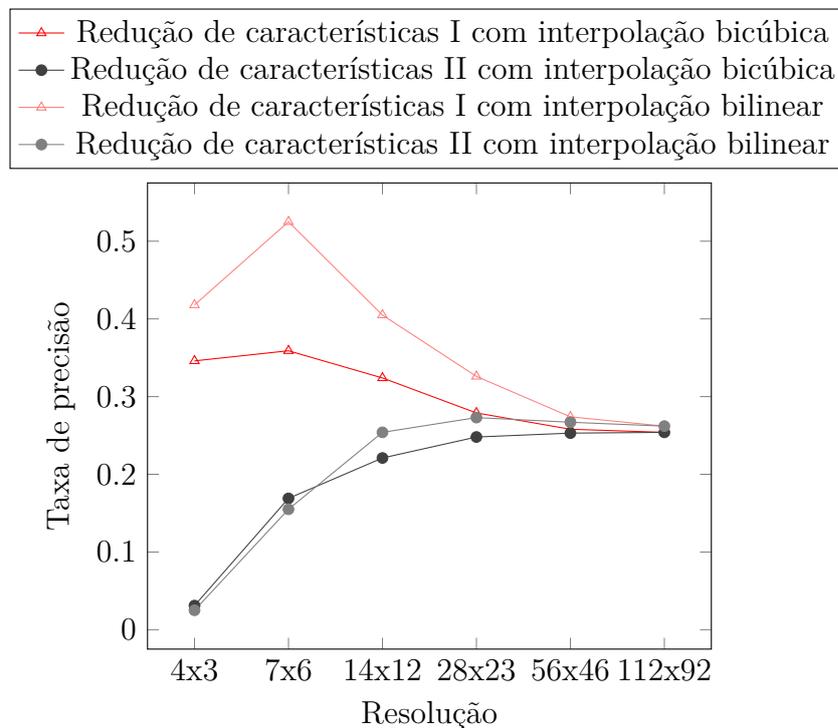


Figura 42 – Precisão por resolução com imagens de treinamento e validação das bases de dados ORL, Yale e Sheffield e imagens de teste das bases de dados ORL, Yale, Sheffield e Georgia Tech

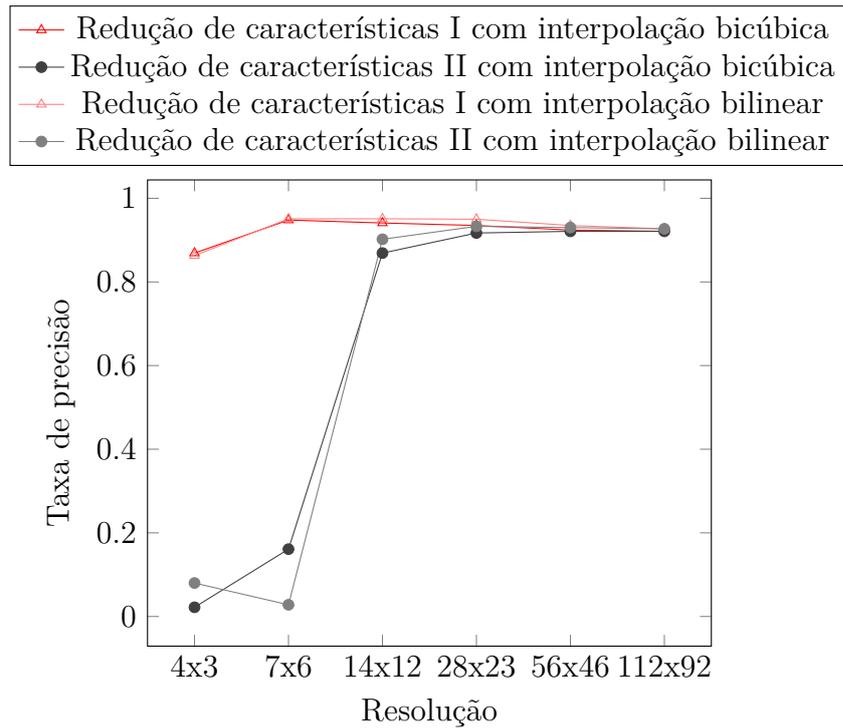


Figura 43 – Precisão Inversa por resolução com imagens de treinamento e validação das bases de dados ORL, Yale e Sheffield e imagens de teste das bases de dados ORL, Yale, Sheffield e Georgia Tech

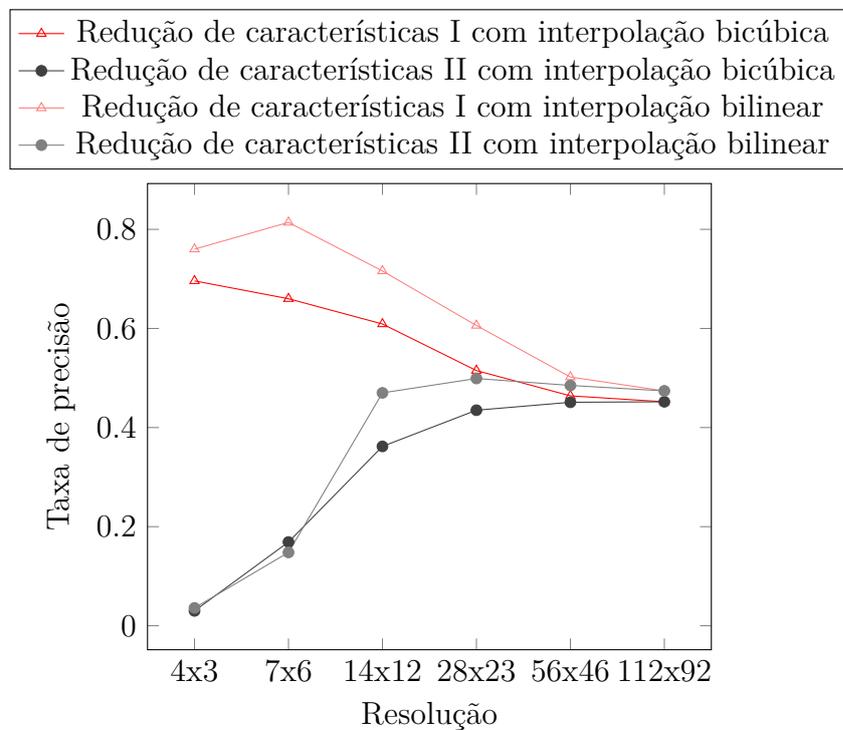


Figura 44 – Acurácia por resolução com imagens de treinamento e validação das bases de dados ORL, Yale e Sheffield e imagens de teste das bases de dados ORL, Yale, Sheffield e Georgia Tech

Capítulo 5

Conclusão

Uma vantagem do reconhecimento facial em relação a outras biometrias é a captura da imagem facial poder se realizada de forma não invasiva, essa característica é importante num contexto em que se está procurando por alguém que não deseja ser identificado, por exemplo, torcedor proibido de frequentar eventos esportivos por possuir histórico violento. A imagem não ser capturada com consentimento causa variações na imagem facial que dificultam o reconhecimento, como posicionamento em relação a câmera, distância para a câmera e oclusão pelo uso de óculos ou chapéu. Particularmente em relação a distância para a câmera, se a pessoa está longe da câmera a região da sua face na imagem é pequena resultando numa imagem facial de baixa resolução. Além da distância entre a pessoa e a câmera, a resolução de captura da câmera também tem influência na obtenção de imagens faciais de baixa resolução.

Alguns métodos são sensíveis a dimensão dos dados de entrada e necessitam que a dimensão dos dados de teste seja a mesma dos dados de treinamento. Normalmente, as imagens usadas para treinamento podem ser capturadas de forma mais controlada numa boa iluminação, pose frontal, expressão facial normal e pequena distância entre câmera e pessoa. As imagens de teste podem ser capturadas num contexto em que não é possível replicar o ambiente das imagens de treinamento. Ao analisar uma imagem facial de teste de resolução diferente da imagem de treinamento é necessário realizar um pré-processamento. Uma abordagem é realizar a análise da imagem na resolução baixa, ou seja, não há uma tentativa de melhoramento da imagem. As imagens de alta resolução são redimensionadas ou a extração de características é invariante a escala. Outra abordagem é aumentar a resolução da imagem para então analisá-la, o redimensionamento pode ser feito por métodos de interpolação ou por métodos mais complexos de melhoramento de imagens. O tratamento do problema de resolução pode ser realizado no nível de classificação. A combinação de classificadores é uma opção usada, os classificadores não precisam tratar de forma especial as imagens de baixa resolução, espera-se que a maioria dos classificadores da combinação não seja afetada minimizando a queda desempenho.

Os métodos propostos nesse trabalho são executados na fase de extração de características. Para realizar a extração de características da imagem de teste com resolução menor que a imagem de treinamento, os vetores de características extraídos das imagens de treinamento são redimensionados para resolução da imagem de teste. A Redução de Características I aplica o redimensionamento nas imagens de treinamento e então as projeta no espaço de características pelos vetores de características reduzidos. A Redução de Características II multiplica os vetores de características reduzidos por um fator de compensação proporcional a taxa de redução da dimensão dos vetores. Nos dois métodos não há necessidade de retreinar o sistema, já que as mesmas características usadas na projeção das imagens são redimensionadas para resolução da imagem de teste. A Redução de Características II tem uma arquitetura mais simples, pois não é necessário redimensionar também as imagens.

O desempenho de Redução de Características I foi superior ao desempenho de Redução de características II. O desempenho da Redução de Características II teve seu desempenho prejudicado quando grandes reduções de fator $\frac{1}{16}$ e $\frac{1}{32}$ foram aplicadas nos vetores de características. Em reduções intermediárias de fator $\frac{1}{2}$, $\frac{1}{4}$ e $\frac{1}{8}$ em alguma bases de dados os resultados foram bem semelhantes. Em comparação com redução de treinamento e *Eigenface*, Redução de Características I foi superior aos dois métodos tendo o desempenho semelhante a redução de treinamento. A maior vantagem de Redução de Características I ocorre nas baixas taxas de erro quando é testado com imagens que não estão na base de treinamento.

5.1 Trabalhos futuros

Redução de características pode ser aplicado em vetores de características de outras técnicas variantes semelhantes ao *Eigenface* que foram citadas neste trabalho 2DPCA, Diagonal PCA, Modular PCA e Kernel PCA. A classificação nos experimentos foi feita pelo KNN, em trabalhos futuros podem ser adicionados métodos mais elaborados de classificação ou até mesmo ter alguns KNN após os métodos de extração de características discutidos e usar um método de combinação de classificadores.

Referências

- ANGGRAINI, D. Face recognition using principal component analysis and self organizing maps. In: *Student Project Conference (ICT-ISPC), 2014 Third ICT International*. [S.l.: s.n.], 2014. p. 91–94.
- BAKER, S.; KANADE, T. Hallucinating faces. In: *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. [S.l.: s.n.], 2000. p. 83–88.
- BEALE, R.; JACKSON, T. *Neural Computing - An Introduction*. [S.l.]: CRC Press, 1990.
- BELHUMEUR, P. N.; HESPANHA, J. P.; KRIEGMAN, D. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, IEEE*, v. 19, n. 7, p. 711–720, 1997.
- BERTILLON, A. *Signaletic Instructions: Including the Theory and Practice of Anthropometrical Identification ... Trans. from the Latest Fr. Ed.* [S.l.]: Werner Company, 1896.
- CAMPBELL J.P., J. Speaker recognition: a tutorial. *Proceedings of the IEEE*, v. 85, n. 9, p. 1437–1462, Sep 1997.
- CAVALCANTI, G. D.; REN, T. I.; PEREIRA, J. F. Weighted modular image principal component analysis for face recognition. *Expert Systems with Applications*, Elsevier, v. 40, n. 12, p. 4971–4977, 2013.
- CHELLAPPA, R.; SINHA, P.; PHILLIPS, P. Face recognition by computers and humans. *Computer*, v. 43, n. 2, p. 46–55, Feb 2010.
- DAUGMAN, J. How iris recognition works. *Circuits and Systems for Video Technology, IEEE Transactions on*, v. 14, n. 1, p. 21–30, Jan 2004.
- DELAC, K.; GRGIC, M. A survey of biometric recognition methods. In: *Electronics in Marine, 2004. Proceedings Elmar 2004. 46th International Symposium*. [S.l.: s.n.], 2004. p. 184–193.
- DUNTEMAN, G. H. *Principal Components Analysis*. [S.l.]: SAGE Publications, 1989.
- EBRAHIMPOUR, R. et al. Low resolution face recognition using combination of diverse classifiers. In: *IEEE. Soft Computing and Pattern Recognition (SoCPaR), 2010 International Conference of*. [S.l.], 2010. p. 265–268.

- EBRAHIMPOUR, R. et al. Low resolution face recognition using mixture of experts with different representations. In: IEEE. *Soft Computing and Pattern Recognition (SoCPaR), 2011 International Conference of*. [S.l.], 2011. p. 475–480.
- ERTAŞ, F. Fundamentals of speaker recognition. *Pamukkale University Journal of Engineering Sciences*, v. 6, n. 2-3, 2011.
- FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, Elsevier, v. 55, n. 1, p. 119–139, 1997.
- GOLDSTEIN, A. J.; HARMON, L. D.; LESK, A. B. Identification of human faces. *Proceedings of the IEEE*, IEEE, v. 59, n. 5, p. 748–760, 1971.
- GONZALEZ, R.; WOODS, R. *Processamento de imagens digitais*. [S.l.]: Edgard Blucher, 2000.
- GORDON, G. G. Face recognition based on depth maps and surface curvature. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *San Diego, '91, San Diego, CA*. [S.l.], 1991. p. 234–247.
- GOTTUMUKKAL, R.; ASARI, V. K. An improved face recognition technique based on modular {PCA} approach. *Pattern Recognition Letters*, v. 25, n. 4, p. 429 – 436, 2004.
- HARDIE, R.; BARNARD, K.; ARMSTRONG, E. Joint map registration and high-resolution image estimation using a sequence of undersampled images. *Image Processing, IEEE Transactions on*, v. 6, n. 12, p. 1621–1633, Dec 1997.
- INDUMATHI, R.; PALANIVEL, N. Pose invariant face recognition using hmm and svm using pca for dimensionality reduction. In: *Information Communication and Embedded Systems (ICICES), 2014 International Conference on*. [S.l.: s.n.], 2014. p. 1–6.
- JACOBS, R. A. et al. Adaptive mixtures of local experts. *Neural computation*, MIT Press, v. 3, n. 1, p. 79–87, 1991.
- JAIN, A. et al. An identity-authentication system using fingerprints. *Proceedings of the IEEE*, v. 85, n. 9, p. 1365–1388, Sep 1997.
- JIANG, L. et al. Survey of improving k-nearest-neighbor for classification. In: *FSKD (1)*. [S.l.: s.n.], 2007. p. 679–683.
- JOSE, J.; POORNIMA, P.; KUMAR, K. A novel method for color face recognition using knn classifier. In: *Computing, Communication and Applications (ICCCA), 2012 International Conference on*. [S.l.: s.n.], 2012. p. 1–3.
- KAUFMAN, G. J.; BREEDING, K. J. The automatic recognition of human faces from profile silhouettes. *Systems, Man and Cybernetics, IEEE Transactions on*, IEEE, n. 2, p. 113–121, 1976.
- KAYA, Y.; KOBAYASHI, K. A basic study on human face recognition. *Frontiers of pattern recognition*, New York: Academic Press, v. 2652289, 1972.
- KIM, K. I.; JUNG, K.; KIM, H. J. Face recognition using kernel principal component analysis. *Signal Processing Letters, IEEE*, v. 9, n. 2, p. 40–42, Feb 2002.

- KUNCHEVA, L. I. *Combining pattern classifiers: methods and algorithms*. [S.l.]: John Wiley & Sons, 2004.
- LEO, M. J.; MANIMEGALAI, D. 3d modeling of human faces-a survey. In: IEEE. *Trends in Information Sciences and Computing (TISC), 2011 3rd International Conference on*. [S.l.], 2011. p. 40–45.
- LI, C. et al. A statistical pca method for face recognition. In: *Intelligent Information Technology Application, 2008. IITA '08. Second International Symposium on*. [S.l.: s.n.], 2008. v. 3, p. 376–380.
- MAHMUD, F. et al. Human face recognition using pca based genetic algorithm. In: *Electrical Engineering and Information Communication Technology (ICEEICT), 2014 International Conference on*. [S.l.: s.n.], 2014. p. 1–5.
- MITCHELL, T. M. *Machine Learning*. 1. ed. New York, NY, USA: McGraw-Hill, Inc., 1997.
- MORAES, A. F. D. Método para avaliação da tecnologia biométrica na segurança de aeroportos. Escola Politécnica de São Paulo, 2006.
- PARVATHY, S.; NAVEEN, S.; MONI, R. A novel approach for multimodal face recognition system based on modular pca. In: *Computational Systems and Communications (ICCSC), 2014 First International Conference on*. [S.l.: s.n.], 2014. p. 127–132.
- POWERS, D. Evaluation: From precision, recall and f-factor to roc, informedness, markedness & correlation (tech. rep.). *Adelaide, Australia*, 2007.
- REYNOLDS, D. A. An overview of automatic speaker recognition. In: *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)(S. 4072-4075)*. [S.l.: s.n.], 2002.
- SAKAI, T.; NAGAO, M.; KANADE, T. *Computer analysis and classification of photographs of human faces*. [S.l.]: Kyoto University, 1972.
- SCHÖLKOPF, B.; SMOLA, A.; MÜLLER, K.-R. Kernel principal component analysis. In: GERSTNER, W. et al. (Ed.). *Artificial Neural Networks - ICANN'97*. [S.l.]: Springer Berlin Heidelberg, 1997, (Lecture Notes in Computer Science, v. 1327). p. 583–588.
- SCHULTZ, R.; STEVENSON, R. Extraction of high-resolution frames from video sequences. *Image Processing, IEEE Transactions on*, v. 5, n. 6, p. 996–1011, Jun 1996.
- SIROVICH, L.; KIRBY, M. Low-dimensional procedure for the characterization of human faces. *JOSA A*, Optical Society of America, v. 4, n. 3, p. 519–524, 1987.
- TAYLOR, W. Machine learning and recognition of faces. *Electronics Letters, IET*, v. 3, n. 9, p. 436–437, 1967.
- TING, K. M.; WITTEN, I. H. Issues in stacked generalization. *J. Artif. Intell. Res. (JAIR)*, v. 10, p. 271–289, 1999.
- TURK, M.; PENTLAND, A. Face recognition using eigenfaces. In: *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*. [S.l.: s.n.], 1991. p. 586–591.

- VIOLA, P.; JONES, M. Robust real-time face detection. *International Journal of Computer Vision*, Kluwer Academic Publishers, v. 57, n. 2, p. 137–154, 2004.
- WANG, C. et al. Color face recognition based on 2dpca. In: *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. [S.l.: s.n.], 2008. p. 1–4.
- WANG, J.; ZHANG, C.; SHUM, H.-Y. Face image resolution versus face recognition performance based on two global methods. In: CITESEER. *Asian Conference on Computer Vision. Jeju Island, Korea*. [S.l.], 2004.
- YANG, J.; YANG, J. yu. From image vector to matrix: a straightforward image projection technique-impca vs. {PCA}. *Pattern Recognition*, v. 35, n. 9, p. 1997 – 1999, 2002.
- YANG, J. et al. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 26, n. 1, p. 131–137, Jan 2004.
- ZHANG, D.; ZHOU, Z.-H.; CHEN, S. Diagonal principal component analysis for face recognition. *Pattern Recognition*, v. 39, n. 1, p. 140 – 142, 2006.
- ZHAO, W. et al. Face recognition: A literature survey. *ACM Comput. Surv.*, ACM, New York, NY, USA, v. 35, n. 4, p. 399–458, dez. 2003.