

RECONHECIMENTO DE FACES EM POSES
ARBITRÁRIAS POR MEIO DE DESCRITORES
LOCAIS

BRUNO AUGUSTO TEIXEIRA

**RECONHECIMENTO DE FACES EM POSES
ARBITRÁRIAS POR MEIO DE DESCRITORES
LOCAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais - Departamento de Ciência da Computação como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

ORIENTADOR: MARIO FERNANDO MONTENEGRO

Belo Horizonte
Setembro de 2014

© 2014, Bruno Augusto Teixeira.
Todos os direitos reservados.

Teixeira, Bruno Augusto

T266r Reconhecimento de Faces em Poses Arbitrárias por
meio de Descritores Locais / Bruno Augusto Teixeira.
— Belo Horizonte, 2014
xxiv, 54 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de
Minas Gerais - Departamento de Ciência da
Computação

Orientador: Mario Fernando Montenegro

Coorientador(es): Alexei Manso Correa Machado

William Robson Schwartz

1. Computação - Teses. 2. Visão por Computador -
Teses. 3. Sistemas de Reconhecimento de Padrões -
Teses. I. Orientador.II. Coorientador.III. Coorientador
IV. Título.

CDU 519.6*82.10(043)




UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO


FOLHA DE APROVAÇÃO


Reconhecimento de faces com poses arbitrárias em imagens por meio de
descritores locais

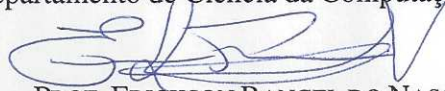
BRUNO AUGUSTO TEIXEIRA

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:


PROF. MARIO FERNANDO MONTENEGRO CAMPOS - Orientador
Departamento de Ciência da Computação - UFMG


PROF. WILLIAM ROBSON SCHWARTZ - Coorientador
Departamento de Ciência da Computação - UFMG


PROF. ALEXEI MANSO CORREA MACHADO - Coorientador
Departamento de Ciência da Computação - PUC/MG


PROF. ERICKSON RANGEL DO NASCIMENTO
Departamento de Ciência da Computação - UFMG


PROF. FLÁVIO LUIS CARDEAL PÁDUA
Departamento de Computação - CEFET/MG

Belo Horizonte, 10 de dezembro de 2014.

Às minhas mães e aos meus irmãos.

Agradecimentos

Após esses 30 meses de trabalho e muito esforço, seria difícil agradecer nominalmente a todas as pessoas que me apoiaram e que contribuíram direta e indiretamente para o desenvolvimento dessa dissertação e para o meu crescimento. De antemão, gostaria de me desculpar com as pessoas que, por ventura, eu tenha esquecido neste texto. Assim, meu primeiro agradecimento é ao meu Deus e Pai, por todas as coisas boas que vivi. Minha felicidade eu só devo à ele, que me guia e me guiou durante toda essa trajetória.

Agradeço também o privilégio que tive de ter sido orientado pelo Prof. Mário Campos, no qual foi paciente e compreensivo durante toda esta etapa, não há como expressar minha gratidão a um dos melhores professores que já tive. Agradeço, sobretudo, os conselhos de pai para filho que recebi. Ao Prof. Alexei Manso, que me orienta há quase seis anos e que contribuiu muito para o meu interesse na área acadêmica. Ele se tornou mais que um orientador, mas um amigo. Sua dedicação foi primordial em toda minha formação acadêmica. Agradeço ao Prof. William, pelas dicas e sugestões extremamente relevantes para o desenvolvimento deste trabalho. Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro para o desenvolvimento desta dissertação. Aos colegas da Accenture e aos meus familiares e amigos pelo apoio e compreensão durante esta jornada.

Por fim, "Agradeço à todas as dificuldades que enfrentei; não fosse por elas, eu não teria saído do lugar. As facilidades nos impedem de caminhar."

A todos vocês o meu muito obrigado.

“Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito.

Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes.”

(Marthin Luther King)

Resumo

O Reconhecimento de Faces de forma confiável ainda constitui um desafio para os atuais sistemas automáticos que desempenham essa tarefa. Imagens adquiridas de indivíduos em ambientes não controlados são fortemente influenciadas por fatores como variação na iluminação, mudança de pose e oclusão. Dentre esses fatores, a variação de pose é o que traz maior impacto ao desempenho dos algoritmos clássicos, pois esses assumem que as características da face não estejam parcial ou completamente ocultas pela rotação. Neste trabalho, é proposta uma abordagem baseada em modelos canônicos tridimensionais da cabeça e em descritores locais para o reconhecimento de faces em poses arbitrárias. A abordagem desenvolvida é composta por dois módulos principais, sendo que o primeiro módulo é responsável por eliminar a limitação em se ter apenas uma imagem de um indivíduo para o reconhecimento. Nesse módulo, uma imagem frontal é recebida como entrada e imagens sintéticas da face são geradas em múltiplas poses. O segundo módulo, por sua vez, consiste no processo de validação da abordagem proposta. Esse módulo basicamente recebe uma imagem de teste como entrada na qual a face está rotacionada num ângulo qualquer, seu ângulo é inferido e os descritores são extraídos para o reconhecimento. A técnica desenvolvida nesta dissertação foi avaliada sob duas vertentes do processo de reconhecimento: a Identificação e a Verificação. Resultados mostram que tal abordagem pode ser integrada aos atuais sistemas de reconhecimento de faces a fim de auxiliar no reconhecimento de faces com poses extremas.

Abstract

Reliable face recognition still stands as a challenge to current automatic face recognition systems. Images acquired from subjects in uncontrolled environments may be strongly affected by numerous factors such as lighting variation, changes in pose and occlusion. This difficulty is due to a number of factors that occur in the images acquired in uncontrolled environments, such as change in illumination, rotation and occlusion. Among those, pose variation is one of the most difficult problems to be overcome, and classical face recognition algorithms assume that facial features are not partially or completely occluded. This work proposes an approach based on local descriptors. The pose variation has a significant effect on performance of the algorithms because most of approaches are based on features of the face that can be partially or completely hidden by rotation. In this paper, we propose an approach based on canonical three-dimensional models of the head and local descriptors for face recognition under arbitrary poses. The developed approach was evaluated in two areas of the face recognition process: identification and verification. Results show that this approach can be coupled in the current face recognition systems for successful face recognition in arbitrary poses.

Lista de Figuras

1.1	Exemplos de variações sistemáticas e não-sistemáticas nas características faciais [Lanitis, 2004].	3
2.1	Configuração de um sistema genérico para o processamento de faces. O diagrama representa cada etapa de um sistema de processamento de faces.	7
2.2	Exemplos de um espaço hipotético de faces com a decisão de cada uma das tarefas de Reconhecimento de Faces - Mckenna et al. [1997].	9
2.3	Exemplo do fenômeno que ocorre nas variações interclasse e intraclasse, adaptado de [Li et al., 2012].	9
2.4	Possíveis ângulos de rotação da Face - Murphy-Chutorian & Trivedi [2009].	12
3.1	Fluxograma do módulo gerador das faces sintéticas em diferentes poses a partir de uma imagem.	20
3.2	Exemplos da saída dos pontos fiduciais detectados para a construção do modelo sintético da face.	20
3.3	Representação do Mapa UV de três indivíduos diferentes.	24
3.4	Representação do processo de recuperação da textura de uma imagem para o 3DMM. Yu [2007]	25
3.5	Módulo que realiza o processo de reconhecimento de faces do sistema desenvolvido. Nesse módulo, é realizada a tarefa de verificação das faces. . .	26
3.6	Exemplos de saída do algoritmo proposto por Zhu & Ramanan, no qual são localizados os pontos fiduciais e no qual também é estimado o ângulo de rotação da face. Os pontos fiduciais localizados serão utilizados para a extração das regiões utilizadas para a classificação.	26
3.7	Exemplo do Operador LBP com uma máscara 3x3.	27
4.1	Exemplos de Imagens da Base de Faces FERET.	30
4.2	Exemplos de imagens com variação de ângulo do Banco de Imagens LATIM.	30

4.3	Exemplo de imagem como parâmetro de entrada para geração de imagens sintéticas.	31
4.4	Exemplo de imagem como parâmetro de entrada para geração de imagens sintéticas.	31
4.5	Exemplo da localização dos pontos fiduciais da Figura 4.3, recebida como parâmetro de entrada para geração de imagens sintéticas.	32
4.6	Exemplo a localização dos pontos fiduciais da Figura 4.4, recebida como parâmetro de entrada para geração de imagens sintéticas.	32
4.7	Exemplo das imagens sintéticas geradas a partir da Figura 4.3, recebida como parâmetro de entrada para o primeiro módulo.	32
4.8	Exemplo das imagens sintéticas geradas a partir da Figura 4.4, recebida como parâmetro de entrada para o primeiro módulo.	33
4.9	Exemplos de falha na etapa de detecção dos pontos fiduciais.	33
4.10	Exemplos de falha na etapa de detecção dos pontos fiduciais.	34
4.11	Exemplos de falhas na etapa de detecção dos pontos fiduciais, nos quais os pontos ultrapassaram o contorno da face.	34
4.12	Exemplo do erro propagado na etapa de detecção dos pontos fiduciais, no qual as imagens sintéticas foram geradas de forma errônea pelo fato do detector não ter localizado precisamente os pontos do contorno da face. . .	35
4.13	Exemplos dos pontos fiduciais detectados para a extração dos descritores, no qual a localização espacial dos pontos é discreta.	36
4.14	ROC da taxa de verificação do conjunto de imagens <i>qr</i> do banco FERET para diferentes tamanhos de blocos utilizados no descritor.	37
4.15	Avaliação dos resultados em cada região para o conjunto de imagens <i>fb</i> do banco FERET.	38
4.16	Comparação dos resultados para abordagem holística versus local em imagens da partição <i>hr</i> do banco FERET, categorizadas como grandes rotações.	39
4.17	Comparação dos resultados para as abordagens tradicionais versus a abordagem proposta em faces frontais do banco LATIM.	40
4.18	Comparação dos resultados para as abordagens tradicionais versus a abordagem proposta em faces rotacionadas à 30° do banco LATIM.	41
4.19	Comparação dos resultados para as abordagens tradicionais versus a abordagem proposta em faces rotacionadas à 45° do banco LATIM.	42
4.20	Comparação dos resultados para o reconhecimento de faces em diferentes ângulos à uma taxa de 10% de erro.	42

Lista de Tabelas

4.1	Resultados obtidos para a tarefa de verificação de faces nas diferentes abordagens.	39
4.2	Área sob a curva (AUC) para cada abordagem de cada experimento. . . .	41

Lista de Siglas

<i>Acrônimo</i>	<i>Descrição</i>
ASM	<i>Modelo de Forma Ativo</i>
AU	<i>Ações de Unidade</i>
DOF	<i>Graus de Liberdade</i>
ELBP	<i>Padrão Binário Local Extendido</i>
GFV	<i>Geração de Faces Virtuais</i>
LBP	<i>Padrão Binário Local</i>
LBPH	<i>Histograma Local do Padrão Binário</i>
LDA	<i>Análise de Discriminante Linear</i>
MPCA	<i>Análise de Componentes Principais Modular</i>
PCA	<i>Análise de Componentes Principais</i>
RBF	<i>Função de Base Radial</i>
ROC	<i>Característica de Operação do Receptor</i>
SVM	<i>Máquinas de Vetores de Suporte</i>
TFA	<i>Análise de Fatores Vinculados</i>
2DPCA	<i>Análise de Componentes Principais em duas dimensões</i>
3DMM)	<i>Modelo 3D Deformável</i>

Sumário

Agradecimentos	ix
Resumo	xiii
Abstract	xv
Lista de Figuras	xvii
Lista de Tabelas	xix
Lista de Siglas	xxi
1 Introdução	1
1.1 Motivação	3
1.2 Objetivos	4
1.3 Organização	5
2 Revisão Bibliográfica	7
2.1 Reconhecimento de Faces	8
2.1.1 Abordagens Holísticas	10
2.1.2 Abordagens Locais	11
2.2 Estimação da Pose	12
2.2.1 Estimação de pose por Rastreamento de Faces	13
2.2.2 Estimação de pose por métodos baseados em Modelos 3D	14
2.2.3 Estimação de pose por Aprendizado de Máquina	15
2.2.4 Estimação de pose por Abordagens Híbridas	15
2.3 Reconstrução 3D da Face	15
3 Metodologia	19
3.1 Geração de Faces Virtuais	19

3.1.1	Detecção de Pontos Fiduciais	20
3.1.2	Modelo 3D Deformável	20
3.1.3	Ajuste 3D da Face	22
3.2	Reconhecimento de Faces	25
3.2.1	Estimação da Pose	25
3.2.2	Matching	26
4	Resultados	29
4.1	Banco de Faces	29
4.1.1	FERET	29
4.1.2	LATIM	30
4.2	Procedimento Experimental	30
4.2.1	Geração de Imagens Sintéticas	30
4.2.2	Reconhecimento Facial	34
4.3	Experimentos	35
4.3.1	Variação do tamanho do bloco	35
4.3.2	Reconhecimento por regiões	36
4.3.3	Validação de Grandes Rotações	37
4.3.4	Validação das Abordagens Holísticas	38
4.3.5	Comparação com as Abordagens Holísticas	40
5	Conclusão	43
	Referências Bibliográficas	45
	Apêndice A Algoritmo Ajuste da Face	51
	Apêndice B Fórmulas de Minimização	53

Capítulo 1

Introdução

Sistemas de processamento facial são sistemas biométricos que, na maioria das vezes, baseiam-se em características únicas extraídas da face, as quais são utilizadas como entrada em diversas abordagens de processamento para a classificação. Um dos sistemas de processamento facial em imagens utilizado em diversas aplicações é o Reconhecimento Facial no qual são necessários algoritmos robustos que atuem em diferentes expressões faciais, sob diferentes condições de iluminação e orientações [Jain et al., 2011]. Dentre essas condições, a variação de pose impinge uma maior degradação aos resultados das técnicas de Reconhecimento Facial em comparação à variação de iluminação e à expressão facial. Esse impacto ocorre pelo fato da face de um mesmo indivíduo possuir menor semelhança sob poses diferentes quando comparada a faces de indivíduos diferentes numa mesma pose.

Embora o ser humano possua, desde criança, certa facilidade para reconhecer rapidamente e sem esforço uma pessoa, segundo [Fanelli et al., 2012], o reconhecimento facial por meio das diferentes abordagens propostas nas últimas décadas falha em realizar tal tarefa com precisão, uma vez que tais métodos devem ser tolerantes a fatores como distorções nas imagens, rotação, iluminação de múltiplas fontes não lambertianas, assim como a fatores biológicos, como expressões faciais, e à presença de acessórios que causam oclusão [Fanelli et al., 2012]. Esta dissertação propõe uma nova abordagem para o Reconhecimento de Faces em poses arbitrárias por meio de descritores locais.

Existem diversas abordagens que tratam o problema de reconhecimento de faces sob poses arbitrárias. De acordo com Shahdi & Abu-Bakar [2012], as abordagens são categorizadas sob aspectos que operam com imagens ou modelos 3D e baseadas em características locais ou globais. A principal diferença entre as abordagens 3D e as abordagens 2D está na precisão na localização das informações extraídas da face, que influenciam diretamente na acurácia do reconhecimento facial. Apesar das abordagens

3D oferecerem alta precisão dos resultados, essas são mais complexas e exigem alto custo computacional. Enquanto que as abordagens 2D possuem particularidades que proveem resultados de acordo com a metodologia utilizada para tratar o problema. Nas abordagens holísticas, decisões para o reconhecimento são tomadas utilizando toda a área da face. Em contraste com as abordagens locais, que utilizam um conjunto de pontos ou regiões pré-definidas. Além dessas classificações dos métodos, segundo Zhao et al. [2003], há também abordagens híbridas nas quais são combinadas abordagens de dois aspectos para se alcançar o mesmo objetivo com resultados significativamente melhores.

Um outro processo que está associado aos sistemas automáticos de reconhecimento de faces em múltiplas poses é a estimativa da pose da face. Ainda que possa parecer um conceito implícito para visão computacional, estimar a pose aceita diversas interpretações. Em visão computacional, estimar a pose da face é o processo de inferir a localização e a orientação de uma cabeça humana a partir de imagens digitais ou vídeos. Essa tarefa exige uma série de etapas de processamento para transformar uma representação de baixo nível baseada em pixel em um conceito de alto nível [Murphy-Chutorian & Trivedi, 2009]. As técnicas para esse processamento podem ser categorizadas de acordo com sua abordagem:

Métodos de rastreamento: recuperam informações do movimento da face observada numa sequência de quadros de vídeo.

Métodos geométricos: utilizam a localização de características, tais como os olhos, boca e nariz para determinar pose.

Métodos de aprendizagem: efetuam, a partir de uma série de imagens, um treinamento para classificar as imagens de entrada.

Métodos baseados em modelo de face: comparam uma imagem de entrada com um conjunto de imagens, a fim de encontrar o ponto de vista mais semelhante.

Métodos híbridos: combinam um ou mais dos métodos acima mencionados para superar as limitações inerentes a qualquer abordagem única.

Abordagens para reconstrução facial também são acopladas aos sistemas de processamento facial que, a partir de uma imagem com a face frontal, gera um modelo sintético 3D dessa face de acordo com o escopo da aplicação. Essas técnicas têm sido utilizadas como pré-processamento a fim de melhorar a eficácia de aplicações para reconhecimento facial. No entanto, o acoplamento dessas técnicas podem impactar na performance, caso tais técnicas não sejam agrupadas de forma correta.

1.1 Motivação

O processamento de faces em imagens estáticas e vídeos é de grande interesse para numerosas aplicações, tais como no processamento de vídeos e imagens, em sistemas de identificação criminal, em sistemas de segurança e na identificação e no monitoramento de pessoas. Essas aplicações também são amplamente requeridas em sistemas de autenticação e identificação que são fundamentados em técnicas de reconhecimento facial. Sabe-se que essa tarefa é de fácil realização por nós, humanos, e de difícil execução pelos computadores, pois as aplicações requerem algoritmos robustos para atuarem sobre diferentes fatores que realizam a precisão desses sistemas. Esses fatores podem ser classificados em duas principais categorias: Variação Sistemática e Variação Não-Sistemática [Lanitis, 2004]. As Variações Sistemáticas são alterações na iluminação, rotação, expressão e as Não-Sistemáticas são causadas por oclusão por meio de objetos, conforme exemplificado na Figura 1.1.



Figura 1.1: Exemplos de variações sistemáticas e não-sistemáticas nas características faciais [Lanitis, 2004]. As imagens da primeira linha exemplificam as variações de expressão facial, de rotação e de iluminação, consideradas como variações sistemáticas. As imagens da segunda linha, por sua vez, representam as variações não-sistemáticas que são causadas pela oclusão de objetos.

Vários métodos foram propostos para a localização e reconhecimento de faces. Contudo, há ainda um conjunto de problemas a serem solucionados que são tratados individualmente, dentre eles está o reconhecimento de faces em poses arbitrárias. As rotações possuem um efeito significativo no desempenho dos algoritmos devido ao fato de a maior parte das abordagens basearem-se em informações da face que possam estar parcialmente ou completamente ocultas pela rotação. A motivação principal para o processo de reconhecimento de faces com precisão é a segurança, dado que os grandes centros urbanos ou locais que possuem maior circulação de pessoas dispõem de câmeras de monitoramento das quais a supervisão é realizada por humanos. Desta forma, com o sistema proposto, o volume de informações disponíveis será processado

em tempo hábil e com margem de erro fixa, enquanto que, se tais informações forem analisadas por humanos, nem todo o volume de dados será avaliado e vários fatores podem influenciar na análise realizada.

Um outro fator motivador é a crescente utilização de tecnologias imersivas com o objetivo de aumentar a interatividade com os usuários. Algumas dessas tecnologias utilizam sensores de profundidade, que se tornaram ao mesmo tempo acessíveis e precisos, e que colocam os usuários no ambiente virtual. A informação de profundidade inerente à essa tecnologia pode ser aproveitada nas técnicas de reconhecimento facial para solucionar problemas de escopo de cada abordagem.

Por fim, as aplicações de processamento de faces ainda possuem restrições devido às variações das imagens de entrada o que as torna pouco eficientes. Essa ineficiência é causada pelas restrições impostas aos sistemas nos quais objetos, como por exemplo a face, estão rotacionados. Com desenvolvimento de uma nova abordagem, podemos prover um método com melhor desempenho e eficácia para o reconhecimento de faces sob poses arbitrárias com variação no ângulo *Yaw*, ou seja rotações da face em torno do eixo x . Tal abordagem é composta de um conjunto de técnicas utilizadas como pré-processamento que são combinadas com o método proposto para tornar o Reconhecimento Facial invariante às restrições.

1.2 Objetivos

Este trabalho tem por objetivo propor uma abordagem para o reconhecimento de faces sob rotação do ângulo *Yaw* em imagens. Para atingir esse objetivo, foi desenvolvido um sistema robusto composto por módulos que não possuem dependência entre si. O principal módulo é composto por uma galeria sintética de imagens em várias rotações geradas a partir de uma imagem frontal de um sujeito qualquer da base de imagens. No outro módulo, é realizada a inferência do ângulo de uma face de teste para a recuperação das imagens na galeria que possua a rotação próxima ao ângulo inferido. No último módulo, é realizado o processo de verificação e identificação das faces. Nesse módulo, descritores são extraídos de características locais para a realização do processo de correspondência. Espera-se que essa abordagem possa contribuir com os sistemas em que a face frontal é parâmetro de entrada, de modo que o acoplamento desta abordagem a outros sistemas torne-os mais robustos. Para isso, será presuposto nos testes que todas as faces de entrada sejam adquiridas de ambiente controlado, ou seja, não possuam rotações *pitch* e *roll*, interferências de iluminação e oclusão.

Por fim, para avaliar a eficácia da abordagem proposta, serão analisadas a taxa de

reconhecimento das faces e o tempo de processamento. Sendo que o processo de extração dos descritores será baseado em técnicas como Análise de Componentes Principais (PCA), Análise de Discriminante Linear (LDA), Análise de Componentes Principais em Duas Dimensões (2DPCA) e Histograma Local do Padrão Binário (LBPH).

1.3 Organização

Esta dissertação está organizada em cinco capítulos. No capítulo 1, foi contextualizado o problema e foram apresentados os objetivos que se espera alcançar neste trabalho. No capítulo 2, é apresentada a revisão bibliográfica da proposta pesquisada, tendo como principal foco a revisão do processo de reconhecimento de faces em múltiplas poses. O Capítulo 3 possui a descrição da metodologia a ser utilizada para atingir o objetivo proposto, como também os materiais necessários para a validação do trabalho. O Capítulo 4 possui a descrição e análise dos resultados dos experimentos realizados. Por fim, o Capítulo 5 contém a conclusão e as possibilidades de trabalhos futuros.

Capítulo 2

Revisão Bibliográfica

Um sistema de processamento de faces pode ser definido como um problema de reconhecimento e processamento de padrões de difícil solução devido à sua não linearidade. Tal sistema possui suas etapas ou subproblemas genericamente descritos de acordo com a Figura 2.1.

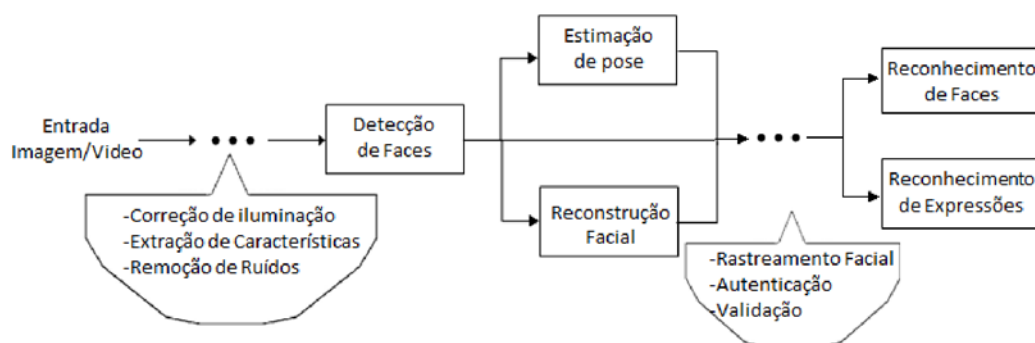


Figura 2.1: Configuração de um sistema genérico para o processamento de faces. O diagrama representa cada etapa de um sistema de processamento de faces.

A face humana contém diversas características que podem ser utilizadas em técnicas de processamento das faces, como olhos, sobrancelhas, boca, nariz. Para a descrição de uma face é necessário analisar o posicionamento, a forma e o tamanho desses elementos, além da distância que separa cada um deles. Essas informações são amplamente utilizadas nos atuais sistemas de reconhecimento de faces, porém nem todos os elementos podem estar presentes na imagem da face para a extração, o que é considerado um problema de difícil solução, uma vez que a busca é feita entre os elementos pertencentes à mesma classe. A maioria dos sistemas não dispõe de várias imagens para treinamento e, quando as imagens são adquiridas sob condições não controladas essas

podem aumentar a complexidade das restrições que os sistemas possuem [Abate et al., 2007].

Sistemas de reconhecimento de padrões baseados em aprendizagem de máquina têm como principal problema, segundo Tan et al. [2006], a utilização de apenas uma imagem para representar cada classe. Este problema é inerente da dificuldade em discriminar classes diferentes, na qual as imagens que representam cada classe possuem similaridade entre si. Para uma melhor classificação nesses sistemas, as imagens de entrada do treinamento devem possuir uma configuração padronizada que não é obtida na aquisição das imagens no mundo real. A seguir, serão descritas as principais abordagens que desempenham cada etapa de um sistema de processamento de faces.

2.1 Reconhecimento de Faces

O Reconhecimento de Faces é composto essencialmente por, no mínimo, quatro tarefas: Identificação, Verificação, Conhecido/Desconhecido e Reconhecimento Completo, [Gong et al., 2000]. Assim, seja uma base de imagens com faces de n pessoas conhecidas, sendo Ω o conjunto de treinamento de todas as classes (ou pessoas) existentes e cada classe representada por $\omega_1, \omega_2, \dots, \omega_i$. A identificação consiste em informar a qual classe $\omega_i \in \Omega$ uma face de teste x pertence, assumindo que essa pessoa está no conjunto, enquanto que a verificação apenas informa se uma face de teste x pode ser classificada numa classe $\omega_i \in \Omega$. A tarefa de Conhecido/Desconhecido visa verificar se uma face x pertence ou não ao conjunto Ω de treinamento. Por fim, o Reconhecimento Completo realiza as três tarefas anteriores, ou seja, dada uma face de teste x , é verificado se esta face pertence ao conjunto Ω de treinamento e, em caso positivo, identifica a classe ω_i desta face. A Figura 2.2 exemplifica as possíveis tarefas supracitadas que podem ser realizadas num modelo criado por um classificador.

Diferenças de pose são uma das principais causas de baixa eficácia nos atuais sistemas de reconhecimento de faces. Isso ocorre devido à oclusão dos elementos faciais extraídos nos quais se baseiam os classificadores e as distorções geométricas não-lineares, que podem ocorrer nos elementos extraídos. Além disso, na variação de pose, ocorre um fenômeno no qual as variações interclasse para diferentes poses possuem maior similaridade do que as variações intraclasse para a mesma pose, conforme Figura 2.3. Ante a essas limitações, foram propostas diversas abordagens que, segundo Shahdi & Abu-Bakar [2012], podem ser categorizadas em dois aspectos - Abordagens 2D X Abordagens 3D e Abordagens Holísticas X Abordagens Locais. A seguir, serão definidas as principais abordagens de cada aspecto que tratam o problema.

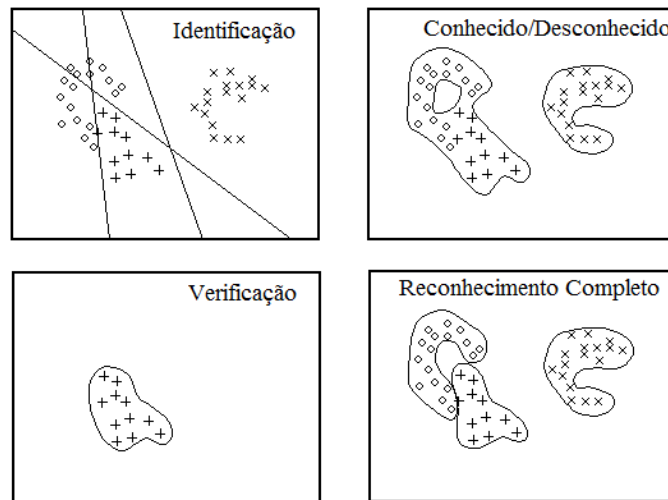


Figura 2.2: Exemplos de um espaço hipotético de faces com a decisão de cada uma das tarefas de Reconhecimento de Faces - Mckenna et al. [1997]. O quadro superior à esquerda possui um conjunto de treinamento com três classes e a tarefa de identificação consiste em informar qual classe uma face teste pertence. Enquanto que o quadro inferior à esquerda exemplifica a tarefa de verificação, na qual uma imagem de teste é avaliada se pode ser classificada numa dada classe de treinamento. O quadro superior à direita exemplifica a tarefa de verificar se um imagem pertence a pelo menos uma das classes de treinamento. O quadro inferior à direita apresenta as três tarefas dos outros quadros.

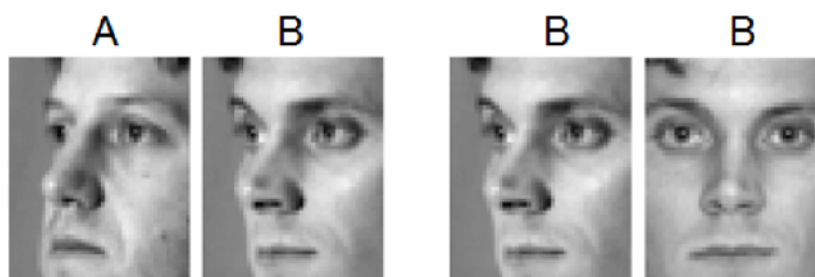


Figura 2.3: Exemplo do fenômeno que ocorre nas variações interclasse e intraclasse, na qual indivíduos diferentes na mesma pose (A e B) possuem maior similaridade comparado imagens do mesmo indivíduo em poses diferentes, adaptado de [Li et al., 2012].

2.1.1 Abordagens Holísticas

As abordagens holísticas têm por princípio a utilização de toda a área da face para a extração de características que serão usadas na construção dos classificadores. Obviamente, essas abordagens oferecem mais informações da face. Todavia, apresentam maior degradação nos resultados devido à incerteza quanto à extração das características, uma vez que essas podem estar ocultas ou sofrerem deformações quando há rotação da face.

Várias técnicas são adotadas para a extração das características que serão usadas na construção dos classificadores. O uso dessas técnicas tem sido um avanço significativo no processo de Reconhecimento Facial [Li & Jain, 2004].

Turk & Pentland [1991b,a] utilizaram Análise de Componentes Principais (PCA) para extrair um conjunto de vetores que são compostos pelos coeficientes mais significantes que representam a face. Tais vetores não necessariamente correspondem às características isoladas, como olhos, orelhas e nariz. Apesar do ganho de desempenho, a técnica não é invariante à escala e à rotação da face. Diante dessas restrições, Demir et al. [2000] propuseram uma abordagem que utiliza várias faces sob diferentes rotações de uma mesma pessoa para o treinamento. Essa abordagem obteve uma taxa de reconhecimento melhor, porém necessita de várias imagens da mesma pessoa para o treinamento, o que a torna limitada. Em [Blanz & Vetter, 2003], o PCA foi aplicado à forma e à textura da face 3D para a extração dos coeficientes que representam a face. Os resultados apresentam alta acurácia, no entanto, a abordagem é computacionalmente cara pelo fato de ser baseada totalmente em modelos 3D, quando aplicado em sistemas embarcados.

Outra abordagem para o reconhecimento facial, proposta por Gao & Leung [2002], é a utilização das informações da borda da face para construção de um mapa que a codifica. Essa técnica obteve melhor taxa de reconhecimento de faces em ambientes não controlados, uma vez que utiliza a medida de similaridade de *Hausdorff*, na qual são avaliadas informações espaciais, estruturais e as orientações das bordas. Esses três componentes na medida de similaridade tornam o método invariante à iluminação. No entanto, segundo Zhang & Gao [2009], essa abordagem é sensível à variação da pose acima de 20° , pois as rotações acima desse ângulo sempre causam distorções no mapa de bordas da imagem.

Em Guo et al. [2011] foi proposta uma abordagem para verificação da face na galeria de imagens utilizando vários descritores extraídos da face. Esses descritores foram aplicados na técnica de regressão *Partial Least Squares* (PLS) e os resultados mostraram que a abordagem é robusta a variações não controladas, especificamente

para variação da pose, uma vez que os descritores não possuem forte correlação. Uma variação do uso do PLS foi proposta por Sharma & Jacobs [2011], que utilizaram apenas as intensidades dos pixels como os descritores. Os resultados desta abordagem foram superiores à técnica de *Tied Factor Analysis* (TFA) que, segundo Zhang & Gao [2009], é o estado da arte em reconhecimento de faces sob vários ângulos. A TFA é uma abordagem probabilística que se diferencia de outras abordagens por utilizar a análise de fatores para a representação dos modelos faciais. Os modelos construídos são treinados pelo algoritmo de Máxima Verossimilhança e a etapa de reconhecimento das faces é baseada na distância entre os modelos [Prince et al., 2008].

2.1.2 Abordagens Locais

As abordagens locais consideram um conjunto de pontos ou regiões específicas da face para extração das características, levando à redução de informações redundantes. Apesar da eliminação de áreas sem importância, as abordagens locais dependem estritamente da correta localização ou extração das características. Brunelli & Poggio [1993] propuseram a extração automática de quatro características de toda a face e, com base na técnica de casamento de padrões, realizaram a comparação da mesma região por meio da correlação cruzada. Essa abordagem possui degradação quando há variação de iluminação e, sobretudo, variação de pose, pois distorções não lineares nas regiões extraídas dos elementos faciais sempre ocorrerão.

Uma extensão da abordagem de [Turk & Pentland, 1991b,a] foi proposta por Pentland et al. [1994], nela a técnica do PCA foi aplicada em subregiões e em toda a face para o reconhecimento facial. A técnica conhecida como Modular PCA (MPCA) se torna invariante à rotação, pois os efeitos de desalinhamento de certas características são eliminados devido à seleção de regiões que não possuem alta distorção quando a face está rotacionada. Fischer et al. consideram que as características extraídas necessitam de um alinhamento padrão, pois a ausência desse fator influencia na performance das técnicas. Além disso, os autores propuseram uma abordagem na qual, após a seleção das regiões faciais, são realizados o alinhamento e a extração dos descritores para aplicação do PLS no processo de reconhecimento.

Além dos métodos estatísticos para o reconhecimento facial supracitados, há também abordagens que se baseiam na textura da imagem. Em [Ahonen et al., 2006], foi proposta uma abordagem que aplica o método *Local Binary Pattern* (LBP) para o reconhecimento de faces. Essa abordagem possui a vantagem de não necessitar da localização precisa das características, uma vez que é analisado o histograma da região da característica.

2.2 Estimação da Pose

Abordagens para estimar a pose das faces têm sido utilizadas como pré-processamento nos atuais sistemas de reconhecimento de faces, de modo a melhorar suas performances. Esse processos consistem em inferir a orientação da face humana em imagens digitais.

No contexto de Visão Computacional, a estimativa da pose facial é a capacidade de inferir a orientação de uma face em relação a um sistema de coordenadas global. Esse conceito, porém, requer conhecimento dos parâmetros intrínsecos da câmera que desfça a distorção na imagem projetada. De acordo com Murphy-Chutorian & Trivedi [2009], essa inferência está intrinsecamente ligada à estimativa da direção do olhar visual no qual se determina a orientação e a localização da face. A amplitude de movimento da face, comumente conhecida como graus de liberdade (DOF), engloba flexões limitadas para cada um dos eixos de rotação. Partindo desse pressuposto, a face está limitada a três ângulos de pose, que podem ser caracterizados pelos ângulos *pitch*, *roll* e *yaw*, conforme representado na Figura 2.4.

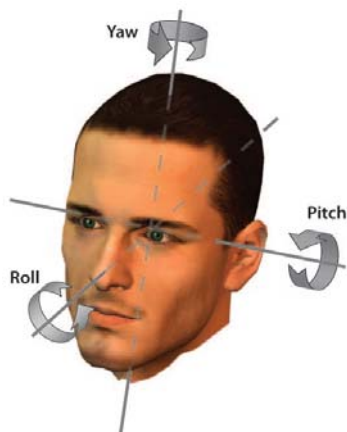


Figura 2.4: Possíveis ângulos de rotação da Face - Murphy-Chutorian & Trivedi [2009].

Diversas abordagens para inferir a pose da face são derivadas de técnicas para estimar o posicionamento de objetos. Simon et al. propõem uma técnica para estimar o posicionamento de objetos em tempo real utilizando sensores de profundidade. Essa abordagem baseia-se na técnica de mínimos quadrados para encontrar pontos do objeto correspondentes em um modelo pré-existente. A utilização deste modelo como particularidade, faz com que o método não necessite *a priori* da extração de características do objeto que será inferido. De acordo Simon et al., essa propriedade elimina a difícil tarefa de extrair características confiáveis de objetos do mundo real. Apesar da abordagem utilizada possuir uma generalização aplicável a qualquer tipo de objeto, a precisão dos ângulos estimados em cada eixo ocorre em função da velocidade de movi-

mento do objeto, por isso a abordagem obtém maior precisão para movimentos lentos e objetos estacionários.

As técnicas para estimar a pose podem ser classificadas, de acordo com a abordagem de implementação utilizada, em técnicas de rastreamento, técnicas baseadas em treinamento, híbridas, entre outras. Uma outra consideração sobre as técnicas é quanto ao seu grau de automatização. Enquanto algumas técnicas estimam a pose automaticamente, outras assumem pré-requisitos que nem sempre podem ser satisfeitos com precisão pelas abordagens disponíveis, tornando essas técnicas limitadas. A seguir, serão definidos os principais métodos utilizados em cada categoria para desempenhar a tarefa de estimação da pose.

2.2.1 Estimação de pose por Rastreamento de Faces

A técnica de rastreamento de faces consiste em realizar a inferência da pose por meio do movimento de rotação da face entre os quadros de um vídeo. As primeiras abordagens propostas por essa categoria utilizam pontos característicos na face para realizar a correlação dos pontos em cada quadro [Gee & Cipolla, 1996]. Métodos que se baseiam na localização de características podem ter uma acurácia altamente instável e baixo desempenho, uma vez que é necessário localizar as características em cada quadro e realizar suas respectivas correspondências através da operação de correlação entre quadros. Abordagens mais sofisticadas [Ozturk et al., 2009] utilizam o SIFT como descritor. Esse por suas características, além de tornar o processo de detecção dos pontos chaves mais robusto, também apresenta um ganho de desempenho no processo de validação, que é baseado na comparação de vetores de pontos das imagens.

Cai et al. [2010] consideram que uma câmera não garante a captura da face para o rastreamento o tempo todo. Por conseguinte, foi proposta uma metodologia para estimar a pose da face pelo rastreamento por meio de múltiplas câmeras, sem a necessidade de informar a pose inicial da face. Embora a abordagem possua essa característica, é necessário realizar a calibração das câmeras e sua aplicação é dispendiosa em sistemas computacionais com recursos limitados.

Em Ba & Odobez [2004], foram utilizados, para estimar a pose, a abordagem de aprendizado de máquina *k-means* juntamente com o método de rastreamento da face para classificar a pose. Nesse trabalho, foram aplicados filtros gaussianos e de Gabor para realizar o treinamento e a classificação das imagens. Desse trabalho, foram obtidos os melhores resultados existentes, porém a técnica não é invariante à rotação *pitch*. Siriteerakul et al. [2011] propuseram uma abordagem baseada na detecção de textura para classificar a pose da face por meio de texturas previamente treinadas.

Nessa abordagem, o rastreamento da face para imagens de baixa resolução obteve resultados significativos como também eficácia para variações na iluminação.

Apesar das abordagens baseadas no rastreamento da face obterem alta precisão do ângulo, muitas das técnicas necessitam da posição inicial da face ou da câmera, que as torna mais complexas em sistemas de execução em tempo real.

2.2.2 Estimação de pose por métodos baseados em Modelos 3D

Métodos para estimar a pose da face por meio de modelos 3D baseiam-se em técnicas para capturar informações de profundidade relativas à face da pessoa presente em uma ou mais imagens. Essas abordagens facilitam o processo de estimativa da posição da face. A despeito disso, a aquisição de modelos 3D baseada na projeção da luz possui restrições quanto às áreas em que não são alcançáveis pelas técnicas e as regiões de baixa ou alta reflexão. Além dessas restrições, técnicas para o registro das imagens adquiridas são necessárias, uma vez que, pequenas variações na representação da imagem para o processo de reconhecimento de faces podem degradar significativamente as taxas de reconhecimento [Araujo & Machado, 2011].

A partir dos modelos em 3D, diversas abordagens são aplicadas para estimar a pose. Malassiotis & Srinivas [2005] consideram que a ponta do nariz é o ponto mais próximo da câmera, diante disso foi proposta uma abordagem baseada na extração dessa característica da face e por meio dela determina-se a curvatura do nariz para, então, estimar sua orientação. A utilização dessa heurística em aplicações do mundo real não é possível devido a suas limitações, na qual é necessário de um ambiente robusto para a construção do modelo 3D em tempo real.

Outra abordagem baseada na utilização de modelos 3D foi proposta em Breitenstein et al. [2008]. Um conjunto de pontos detectados em algumas regiões da face determinam assinaturas das quais se pode determinar a direção da face. Além de ser uma técnica baseada em heurísticas, seu processamento demanda um alto custo computacional.

O acoplamento de outras abordagens para extração de características tem propiciado aplicações mais robustas e com resultados promissores. Em Liang & Tong [2012], a extração da ponta do nariz é feita em conjunto com a segmentação de características como boca e olhos. Essas características são consideradas no processo de classificação da pose realizado pelo conjunto de treinamento de imagens da mesma classe.

2.2.3 Estimação de pose por Aprendizado de Máquina

Diferentes abordagens [Fanelli et al., 2011; Huang et al., 2010, 2011] de aprendizado de máquina também são utilizadas em sistemas para estimar a pose por meio da classificação. [Fanelli et al., 2011; Huang et al., 2010] propõem uma nova abordagem de aprendizado utilizando imagens de profundidade conjugadas com a posição da face no espaço. Mesmo que essas metodologias sejam eficazes, elas ainda são bastantes restritivas quanto à rotação, à mudança de escala e à oclusão nas faces, uma vez que se faz necessário de várias imagens de um mesmo indivíduo sob as diferentes restrições para realizar o aprendizado do modelo [Fanelli et al., 2012].

Em Tang et al. [2011], foi utilizada a teoria de Integral da Imagem em conjunto com Redes Neurais para realizar a classificação da face. Essa abordagem utiliza uma camada do modelo 3D de profundidade para, então, encontrar vetores de direção da face. Os vetores são calculados pela integral de um determinado ponto para realizar a classificação do ângulo estimado. Técnicas baseadas nesse tipo de abordagem devem possuir bases suficientemente grandes disponíveis para que a taxa de acerto seja satisfatória [Fanelli et al., 2012]. Além disso, deve-se ter grande número de imagens para treinamento com biótipos de diferentes pessoas e imagens que não contêm faces. A forma com que o treinamento é conduzido também influencia no resultado final, sobretudo, pela inviabilidade de mudar as condições do treinamento em tempo real, ou seja, as características que foram determinadas para classificação da pose não podem alterar-se a qualquer momento.

2.2.4 Estimação de pose por Abordagens Híbridas

As abordagens híbridas combinam uma ou mais abordagens para estimar pose com o intuito de uma limitação imposta por uma das técnicas ser eliminada pelas outras. Aplicações que suportam esse tipo de abordagem possuem alta precisão e limitações reduzidas, contudo é necessário encontrar pontos comuns entre as técnicas para que não se tenha um alto custo computacional.

2.3 Reconstrução 3D da Face

A geração de um modelo ou reconstrução 3D da Face vem sendo utilizada em diversos sistemas nos quais as aplicações são sensíveis à presença de pessoas. Esse processo de estimar a estrutura 3D da face, segundo He et al. [2008], não pode ser realizado apenas por uma interpolação entre as imagens adquiridas. Blanz & Vetter [1999, 2003]

derivaram um modelo 3D deformável (3DMM) da face utilizando 200 imagens capturadas por um scanner 3D. Esse modelo é considerado computacionalmente eficiente e robusto para a reconstrução da face 3D utilizando uma imagem. O principal dificultador para utilização das técnicas que realizam a reconstrução 3D da face são as regiões em que não são visíveis ou alcançáveis no processo de captura. A abordagem proposta por Blanz & Vetter é considerada uma abordagem de análise por síntese, pelo fato de possibilitar também o reconhecimento de faces a partir dos parâmetros estimados no modelo sintético construído. Existem três principais métodos propostos na literatura considerados como sendo uma abordagem de análise por síntese [Levine & Yu, 2009]. Esses métodos se diferem no modo como serão estimados os parâmetros da forma e da textura do modelo sintético construído. Apesar desses métodos construírem modelos sintéticos com alta precisão, a eficiência é um dos principais gargalos para aplicá-los em sistemas de tempo real. Dada essa limitação, Park et al. [2010] propuseram uma abordagem para estimação dos parâmetros do modelo sintético 3DMM, a partir de um modelo sintético reduzido, tornando-o um método robusto e eficiente para sistemas de reconhecimento facial automático.

A representação da face por modelos sintéticos em malhas de polígonos também é utilizada em diversos trabalhos, não só pela baixa complexidade como também pela redução na precisão de características em regiões da face que não possuem relevância. Em Rydfalk [1987], foi feito um modelo parametrizado da face baseado na descrição da forma facial e nas variações das expressões faciais. As expressões faciais são descritas por ações de unidade (AU), que modificam a expressão facial por meio da ativação de um ou mais músculos. Assim, nesse modelo é possível gerar todas as expressões faciais por meio da combinação das 44 AUs modificando apenas os vértices da malha de polígonos. [Martins, 2008] considera que um modelo antropométrico da face possui melhor descrição da superfície facial de vários indivíduos.

Suen et al. [2007] propuseram uma abordagem baseada em textura da qual, a partir de duas imagens de qualquer ângulo, extraem-se pontos característicos com o intuito de se estimar a pose para, então, aplicar a textura num modelo 3D construído a partir dos pontos. Apesar dessa abordagem não demandar alto custo computacional, é pressuposto que as imagens sejam adquiridas de uma certa distância, como também que o ângulo da face não pode ocultar os elementos faciais extraídos. Tais condições tornam essa abordagem limitada. Ming et al. [2010] consideram que, pelo fato das abordagens não levarem em conta a estrutura da face humana, a maioria delas apresenta pior desempenho quando imagens de entrada são faces de perfil. Em sua proposta, os pontos característicos são extraídos de acordo com presença das características específicas da face como boca, olhos abertos/fechados, entre outras. A avaliação

individual das características da face para extração dos pontos possui a vantagem de extrair características efetivas da face removendo eventuais ruídos.

Diversos métodos baseiam-se na abordagem *shape-from-X* para se recuperar a estrutura tridimensional de um objeto. Nessa abordagem, X representa possíveis variáveis, como textura, e contornos são utilizados para a reconstrução da face. Contudo, esses métodos requerem várias imagens ou sequência de vídeos, e nem sempre é possível se obter várias imagens [Ming et al., 2010].

Capítulo 3

Metodologia

Neste trabalho foi desenvolvido um sistema robusto capaz de realizar automaticamente o reconhecimento de faces em poses rotacionadas no ângulo *Yaw*. A abordagem proposta é composta de dois principais módulos independentes. O primeiro módulo, denominado *Geração de Faces Virtuais (GFV)*, consiste na geração de imagens de faces sintéticas rotacionadas em diferentes ângulos a partir de uma imagem contendo uma face frontal. O segundo módulo, por sua vez, realiza um dos processos de reconhecimento de faces, a verificação, com base nas imagens virtuais geradas. A seguir, serão descritos os principais blocos de cada módulo.

3.1 Geração de Faces Virtuais

O processo de geração de faces virtuais é composto pelas etapas ilustradas na Figura 3.1. Uma imagem com uma face frontal é a entrada para a geração de imagens sintéticas em diferentes ângulos desse módulo. A primeira etapa do GFV é a localização de pontos fiduciais da face, os quais são correspondentes aos pontos definidos, *a priori*, no modelo 3D sintético. Na etapa seguinte, é feita a recuperação da forma 3D e a informação de textura da face. Por fim, a forma 3D e a informação de textura são combinadas para a geração das imagens sintéticas em diferentes ângulos.

A geração das faces sintéticas de forma online aumenta o custo computacional da aplicação. Com o intuito de tornar o sistema eficiente para execução em tempo real, esse módulo foi elaborado para ser executado em *modo batch*, ou seja, em um processo que pode ser executado de forma independente num momento distinto.

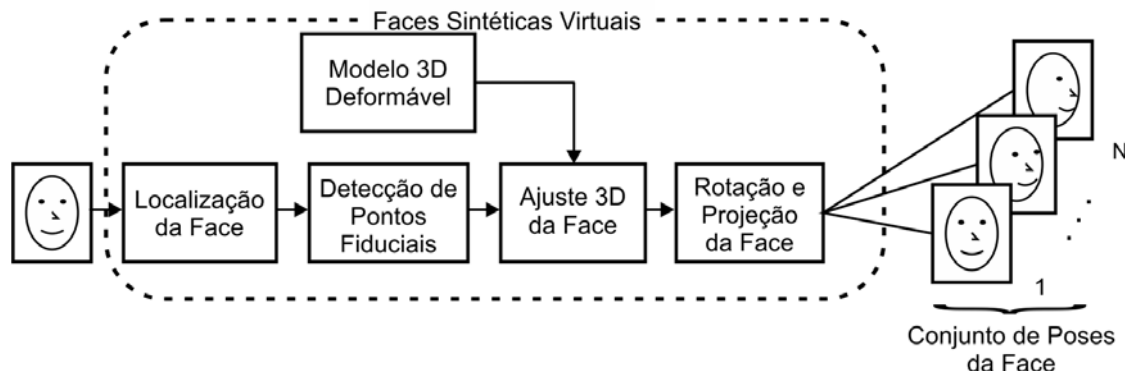


Figura 3.1: Fluxograma do módulo gerador das faces sintéticas em diferentes poses a partir de uma imagem.

3.1.1 Deteção de Pontos Fiduciais

A deteção de pontos fiduciais de forma automática é um processo complexo para muitas aplicações de análise de imagens, pois a imprecisão na localização dos pontos degrada a eficácia dos sistemas. Neste trabalho, foi utilizada a técnica Modelo de Forma Ativo¹ (ASM) para a deteção dos pontos fiduciais [Milborrow & Nicolls, 2008]. A abordagem ASM baseia-se em modelos estatísticos da face que iterativamente vão deformando-se para ajustarem-se à face. A Figura 3.2 exemplifica a localização desses pontos.

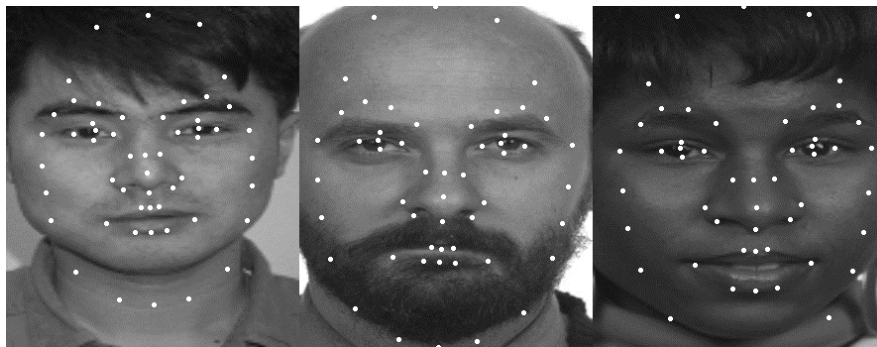


Figura 3.2: Exemplos da saída dos pontos fiduciais detectados para a construção do modelo sintético da face.

3.1.2 Modelo 3D Deformável

Um modelo de face 3D é basicamente composto pela forma e pela textura: a forma é definida pelos vértices do modelo em três dimensões, enquanto que a textura descreve

¹Tradução livre de *Active Shape Model*

as informações de cores. Em Blanz & Vetter [1999], foi construído um Modelo 3D Deformável (3DMM) da face para que qualquer face 3D possa ser gerada a partir de uma combinação linear dos vetores de forma e de textura. Essa combinação linear é controlada pelos parâmetros $\vec{\alpha}$ e $\vec{\beta}$, da forma e da textura, respectivamente, de acordo com as equações:

$$\vec{S} = \sum_{i=1}^m \alpha_i S_i, \quad (3.1)$$

$$\vec{T} = \sum_{i=1}^m \beta_i T_i. \quad (3.2)$$

onde \vec{S} e \vec{T} representam a combinação linear das formas S_i , $S_i \in R^{N \times 3}$, e textura T_i , $T_i \in R^{N \times 3}$, dos m indivíduos utilizados para construir uma face sintética.

Para a construção do 3DMM, os autores utilizaram 100 ($m = 100$) cabeças de humano escaneadas por um laser. Cada cabeça possui um conjunto de N vértices ($N = 75972$), armazenados no vetor da forma da face $S_i = (X_{i,1}, Y_{i,1}, Z_{i,1}, \dots, X_{i,N}, Y_{i,N}, Z_{i,N})^T$ e de textura $T_i = (R_{i,1}, G_{i,1}, B_{i,1}, \dots, R_{i,N}, G_{i,N}, B_{i,N})^T$. Dado que cada vértice do conjunto de vetores das cabeças possui correspondência, Blanz e Vetter aplicaram a análise de componentes principais nos 100 vetores escaneados da cabeça. Assim, a partir do conjunto de vetores $S = [S_1, S_2, \dots, S_m] \in R^{(3 \times N) \times m}$ para a forma e $T = [T_1, T_2, \dots, T_m] \in R^{(3 \times N) \times m}$ para a textura, a forma e a textura média foram obtidas pelas seguintes equações:

$$\vec{S} = \frac{1}{m} \sum_{i=1}^m S_i, \quad (3.3)$$

$$\vec{T} = \frac{1}{m} \sum_{i=1}^m T_i. \quad (3.4)$$

A matriz de covariância da forma, \vec{C}_S , e da textura, \vec{C}_T , é definida pelas equações:

$$\vec{C}_S = \sum_{i=1}^m (S_i - \vec{S})(S_i - \vec{S})^T, \quad (3.5)$$

$$\vec{C}_T = \sum_{i=1}^m (T_i - \vec{T})(T_i - \vec{T})^T. \quad (3.6)$$

Por meio da análise de componentes principais, são extraídos os autovetores e_i^s e e_i^t e os respectivos autovalores λ_i^s e λ_i^t tais que:

$$C_S e_i^s = \lambda_i^s e_i^s, \quad (3.7)$$

$$C_T e_i^t = \lambda_i^t e_i^t. \quad (3.8)$$

Após a construção do 3DMM, é possível recuperar qualquer uma das m cabeças utilizadas para a modelagem com as seguintes equações:

$$\bar{S}_j = \bar{S} + \sum_{i=1}^m \frac{\alpha_i}{\lambda_i^s} e_i^s, \quad (3.9)$$

$$\bar{T}_j = \bar{T} + \sum_{i=1}^m \frac{\beta_i}{\lambda_i^t} e_i^t, \quad (3.10)$$

onde e_i^s e e_i^t representam os i -ésimo autovetores da matriz de covariância. Em adição, Volker Blanz e Tomas Vetter estimaram que os parâmetros $\vec{\alpha}$ e $\vec{\beta}$ seguem uma função de distribuição normal de probabilidade tal que:

$$p_S(\vec{\alpha}) \approx e^{-\frac{1}{2} \sum_{k=1}^m \left(\frac{\alpha_k}{\lambda_k^s}\right)^2}, \quad (3.11)$$

$$p_T(\vec{\beta}) \approx e^{-\frac{1}{2} \sum_{k=1}^m \left(\frac{\beta_k}{\lambda_k^t}\right)^2}. \quad (3.12)$$

3.1.3 Ajuste 3D da Face

Devido à grande quantidade de vértices no 3DMM, neste trabalho foi utilizado um modelo deformável simplificado semelhante à abordagem proposta em Park et al. [2010]. O modelo do 3DMM foi reduzido de forma significativa, se comparado ao modelo original, com o intuito de simplificar o processamento, bem como de melhorar a eficiência na derivação dos parâmetros. Esse novo modelo possui 54 vértices correspondentes aos pontos fiduciais que foram detectados na etapa anterior. Desta forma, seguindo a metodologia proposta por Blanz & Vetter [2003], foi aplicado o PCA no novo modelo cujo a quantidade de vértices foram reduzidas, S_{mm} , para a estratificação da média \bar{S}_{mm} , dos autovalores λ_{lS} e dos autovetores W_{lS} da matriz de covariância do novo modelo. Como visto anteriormente, a forma da face 3D do modelo simplificado pode, então, ser representada como:

$$\bar{S}_\alpha = \bar{S}_{mm} + \sum_{l=1}^L \alpha_l W_l, \quad (3.13)$$

onde α é o parâmetro que controla a forma da face, W é a matriz dos autovetores do modelo e L o número de faces do modelo. Assim, para transformar os pontos fiduciais da imagem detectados na etapa anterior, $S_{2d,i}^j$, nos correspondentes pontos do modelo 3D, S_j , foi utilizado o modelo deformável reduzido para computar a função de minimização 3.14, de modo a reduzir o erro global do ajuste do modelo 3D para uma face baseado nos pontos 2D:

$$E(P, R, t, a, \alpha_{l=1}^L) = \|S_{2d,i}^j - T_{P,R,t,a}(S_\alpha)\|^2, \quad (3.14)$$

onde $T(\cdot)$ é um operador composto por transformações, como Rotação (R), Translação (t), escala (a) e projeção (P), que serão aplicadas no modelo simplificado. $S_{2d,i}^j$ representa o conjunto dos pontos fiduciais que foram localizados na face 2D, conforme descrito na seção 3.1.1.

Segundo Park et al. [2010], a detecção dos pontos de forma manual ou automática não recupera uma localização precisa. Essa imprecisão pode levar à geração de modelos não confiáveis. Destarte, o autor utilizou um termo de regularização de *Tikhonov* para controlar a distância entre o modelo ajustado e o modelo médio. O termo σ é o desvio padrão da energia E calculada em 3.14 determinado empiricamente, que resulta numa nova equação de energia regularizada:

$$E' = E/\sigma^2 + \sum_{l=1}^L \alpha_l^2/\lambda_l. \quad (3.15)$$

O processo de minimização da equação 3.15 para estimar os parâmetros (a , α , R e t) pode ser realizado por diferentes abordagens. Neste trabalho, foi utilizada a metodologia proposta por Park et al. [2010], que alternadamente estima os parâmetros da pose, orientação e posição (R, T), e o parâmetro da forma da face (α) até que a energia residual convirja. Essa abordagem se caracteriza por possuir baixo custo computacional, uma vez que a resolução de uma função não-linear é transformada na resolução de uma função linear. Conforme descrito no Apêndice A, o algoritmo para o ajuste 3D da face inicialmente estima os parâmetros da pose (R, T), com os parâmetros da forma (α) inicializados com 0. No próximo passo, os parâmetros da pose são fixados com os valores encontrados e os parâmetros da forma são estimados. Alternadamente, esses parâmetros são estimados até que o erro residual da energia E convirja. As funções de minimização são estimadas por análise numérica, com o método *levenberg-marquardt*. Por fim, a face 3D é reconstruída com os parâmetros (α) da forma estimados.

3.1.3.1 Recuperação da Textura

Um modelo 3D é composto da forma tridimensional e dos dados de textura. Após o ajuste da forma 3D da face na etapa anterior, resta apenas a recuperação da textura para a geração do modelo 3D da face. A recuperação de forma precisa da textura 3D a partir de uma imagem, é realizada por meio do mapa UV, conforme exemplificado pela Figura 3.3. O mapeamento UV é um processo que distorce uma imagem com a informação de textura para uma malha 3D. O processo inverso é conhecido como Processo de Achatamento², no qual a informação de cor do vértice é armazenada numa coordenada UV.

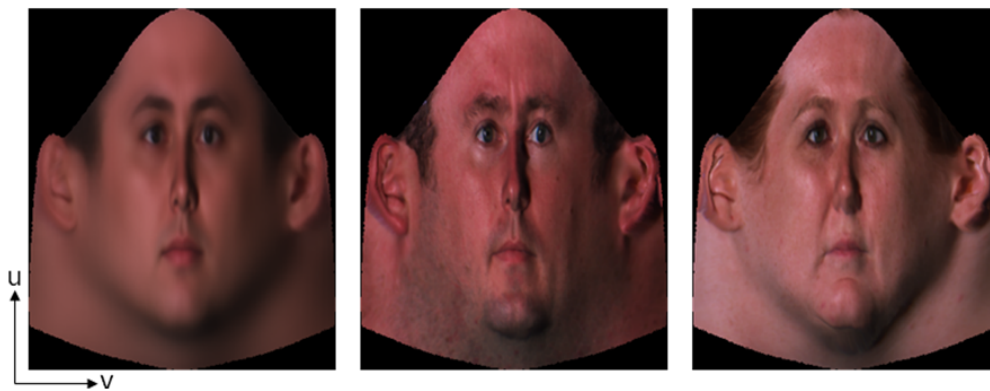


Figura 3.3: Representação do Mapa UV de três indivíduos diferentes. Os três indivíduos possuem a mesma forma facial e diferentes aparências - Yu [2007].

Para realizar a correspondência da forma 3D para o espaço 2D, [Blanz & Vetter, 2003] aplicaram a projeção cilíndrica na imagem para realizar o Processo de Achatamento². O processo de Achatamento é definido por uma função f , que recebe um vértice k de uma forma arbitrária 3D e gera uma única saída (u_k, v_k) no espaço UV. Desta forma, o processo de mapeamento da textura para o conjunto de pontos fiduciais extraídos é realizado de forma eficiente, dado que os pontos possuem correspondência entre si. No entanto, o mapeamento dos pontos não extraídos da imagem para o modelo 3D é uma tarefa complexa, pois apenas o mapa UV possui correspondência direta no modelo 3D. De acordo com a abordagem proposta por [Yu, 2007], a aplicação da textura de uma imagem para o 3DMM pode ser simplificado a um mapeamento 2D-2D envolvendo uma técnica de *warping* que realiza o achatamento para o mapa UV, conforme Figura 3.4.

O mapeamento UV foi realizado por meio de uma Função de Base Radial (RBF) *Thin-Plate Spline* para o *warping*. Os pontos fiduciais da imagem e do modelo serviram

²Tradução livre de *Flattening Process*

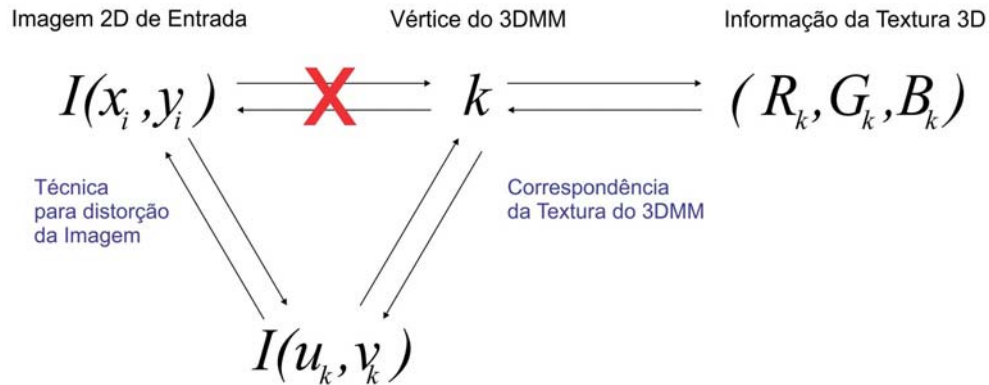


Figura 3.4: Representação do processo de recuperação da textura de uma imagem para o 3DMM. Yu [2007]

de pontos de controle para a rede neural artificial (RNA) do tipo RBF. Esses pontos influenciam na inferência da localização dos pontos que não possuem um mapeamento direto. Após o treinamento da rede neural RBF, os pontos restantes da imagem são deformados para o mapa UV que, por sua vez, é diretamente mapeado no 3DMM.

3.2 Reconhecimento de Faces

O segundo módulo deste trabalho realiza uma das tarefas de reconhecimento das imagens, a verificação, conforme ilustrado na Figura 3.5. A verificação tem por finalidade validar se um par de faces pertence a uma mesma pessoa ou não. Nesse módulo, uma imagem de um indivíduo desconhecido é recebida como entrada para o casamento com as imagens sintéticas geradas pelo módulo anterior. A primeira etapa do módulo localiza os pontos fiduciais e infere o ângulo da face de teste. Dado que o ângulo estimado possui uma margem de erro, foi selecionado um subconjunto das imagens sintéticas para o reconhecimento ao invés de validar apenas com imagens que possuem o mesmo ângulo inferido na face de teste. A partir dos pontos fiduciais localizados, é realizada a extração das características em cada ponto para o casamento com as imagens de teste.

3.2.1 Estimação da Pose

Na etapa de estimação da pose, foi utilizado o método MTSPM proposto por Zhu & Ramanan [2012]. Esse método realiza a estimação do ângulo e também encontra um conjunto de pontos fiduciais da face que serão utilizados para extração das características na etapa de classificação das faces. A imagem 3.6 exemplifica a localização dos pontos fiduciais nas imagens sintéticas.

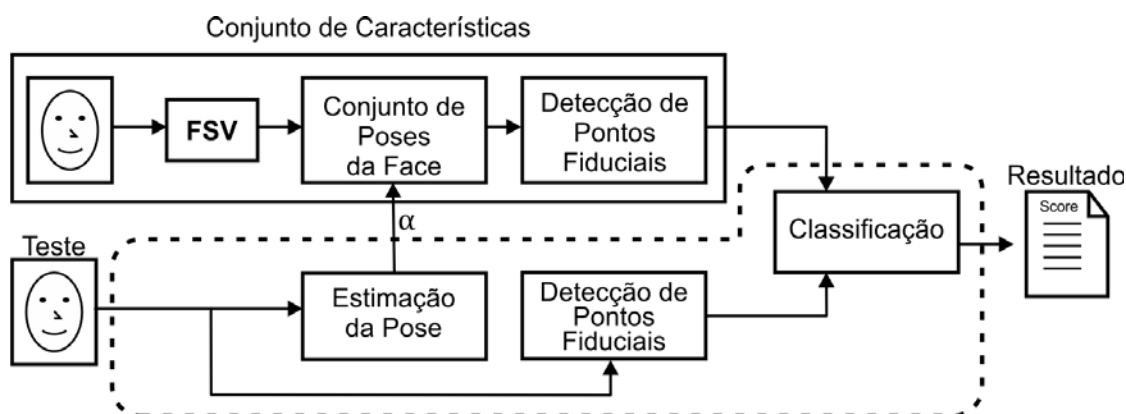


Figura 3.5: Módulo que realiza o processo de reconhecimento de faces do sistema desenvolvido. Nesse módulo, é realizada a tarefa de verificação das faces.

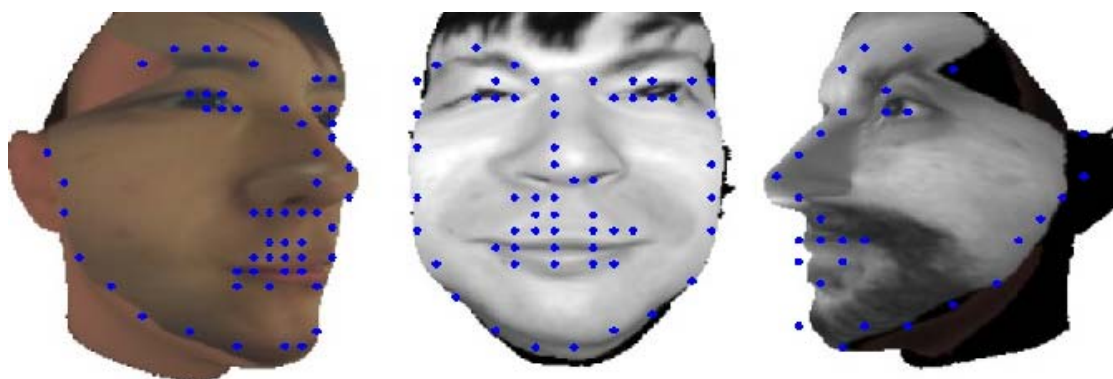


Figura 3.6: Exemplos de saída do algoritmo proposto por Zhu & Ramanan, no qual são localizados os pontos fiduciais e no qual também é estimado o ângulo de rotação da face. Os pontos fiduciais localizados serão utilizados para a extração das regiões utilizadas para a classificação.

3.2.2 Matching

A etapa de matching baseia-se em descritores locais extraídos em cada ponto fiducial detectado pela abordagem supracitada. Esses descritores foram extraídos com o método Padrão Binário Local ³ (LBP) proposto por Pietikäinen et al. [2011], que é um operador não paramétrico que sumariza de forma eficiente estruturas locais da imagem em rótulos inteiros. O operador LBP procede conforme ilustrado na Figura 3.7, em que cada pixel é subtraído dos seus 8 vizinhos de uma máscara 3 x 3. Com o resultado negativo, o valor do pixel na máscara é atribuído 0 ou 1, caso o resultado seja positivo. Dessa forma, um número binário é obtido para cada pixel com concatenação de todos os valores binários da máscara. O valor decimal correspondente do número binário gerado é, então, utilizado para rotular o pixel determinado.

³Tradução livre de *Local Binary Pattern*

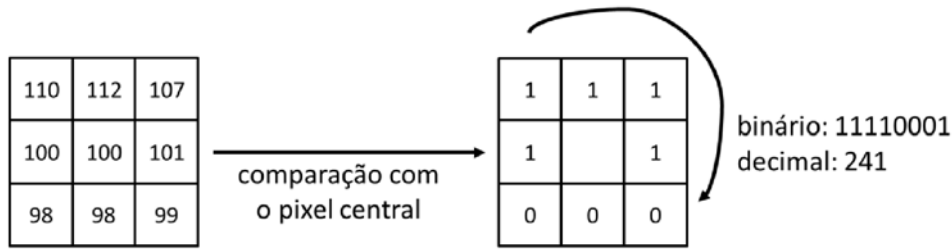


Figura 3.7: Exemplo do Operador LBP com uma máscara 3x3.

Embora o LBP original seja eficiente e invariante à rotação, o tamanho da máscara o torna limitado quanto à descrição de padrões em escalas diferentes. Apesar dessas limitações, Ojala et al. [2002] estenderam esse operador para suportar vizinhanças de pixels com tamanhos, raios e formas variadas, permitindo a descrição de máscaras em diferentes escalas. Portanto, neste trabalho, foi utilizado o operador genérico do LBP, conhecido como ELBP, descrito pela equação 3.16 para um dado pixel (x_c, y_c) :

$$LBP_{P,R}(x_c, y_c) = \sum_{P=0}^{P-1} s(i_P - i_c) 2^P, \quad (3.16)$$

onde i_c e i_P são, respectivamente, os níveis de cinza do pixel central e do pixel adjacente do círculo de raio R e, a função $s(\cdot)$ é definida pela equação 3.17:

$$s(x) = \begin{cases} 1, & \text{se } x \geq 0, \\ 0, & \text{se } x < 0. \end{cases} \quad (3.17)$$

O processamento do ELBP é realizado para todos os pixels da imagem e posteriormente é gerado o histograma para a etapa de classificação. A abordagem proposta em Ahonen et al. [2006] divide toda a imagem em subregiões e aplica o operador em cada uma delas. Em seguida os histogramas dessas subregiões são concatenados para realizar a classificação. Diferentemente dessa abordagem, neste trabalho, a comparação das imagens é realizada por cada região de cada ponto fiducial encontrado nas duas imagens. Essa modificação torna a correspondência espacial da classificação mais flexível, uma vez que as características locais de uma imagem podem ser comparadas às características locais encontradas em posições diferentes de outras imagens.

Para realizar a etapa de *matching*, foi utilizada a métrica de distância do vizinho mais próximo conhecida como χ^2 . Assim, dados dois histogramas extraídos das regiões, a equação:

$$\chi^2(x, y) = \sum_{i=1}^D \frac{(x_i - y_i)^2}{(x_i + y_i)}, \quad (3.18)$$

onde D , representa a dimensão do histograma, resulta o quão parecidas as faces são.

Capítulo 4

Resultados

Conforme descrito anteriormente, este trabalho tem por objetivo realizar o reconhecimento de faces em poses arbitrárias por meio de descritores locais. A partir deste princípio, no qual os descritores locais são utilizados para a etapa de classificação e baseando-se em outros trabalhos propostos, neste capítulo, serão apresentados cinco cenários de testes para validação da abordagem proposta. Foram utilizados dois bancos de faces públicos para a execução de cada categoria dos testes. As subseções a seguir descrevem as características de cada banco de dados e a finalidade de cada cenário de teste, como também os resultados de cada etapa desenvolvida.

4.1 Banco de Faces

4.1.1 FERET

A base de dados FERET [Phillips et al., 1998] contém imagens de 994 pessoas, sendo que cada pessoa possui até 13 imagens capturadas. Essas imagens foram capturadas em laboratório sob condições controladas de luminosidade e plano de fundo. As fotos de algumas pessoas foram tiradas em momentos distintos, porém, devido ao pouco intervalo de tempo, seus efeitos não são perceptíveis. Variações de pose e a presença ou não de óculos transparentes são as características das imagens do banco, que foram subdivididas em vários conjuntos de acordo com o ângulo da face representado pelas partições do protocolo FERET, $(fa, ql, qr, hl, hr, pl, pr)$. Neste trabalho, para todos os experimentos do banco FERET, serão utilizadas apenas as imagens frontais, pertencentes ao conjunto fa , para a construção das imagens sintéticas, e as partições com as imagens sob rotação, para validação dos resultados, conforme ilustrado na imagem 4.1. Todas as imagens utilizadas nos experimentos foram normalizadas na resolução de 296

x 347 pixels.



Figura 4.1: Exemplos de Imagens da Base de Faces FERET.

4.1.2 LATIM

O banco de imagens LATIM é composto de 1120 imagens, adquiridas entre abril de 2010 e junho de 2010, sendo 560 de faces de homens e 560 faces de mulheres. Tais faces foram submetidas a variações de expressão, iluminação, ângulo e oclusão em fundo branco, sendo que a variação de iluminação foi feita em conjunto com as variações de oclusão.



Figura 4.2: Exemplos de imagens com variação de ângulo do Banco de Imagens LATIM.

4.2 Procedimento Experimental

O processo de validação da metodologia proposta consiste em avaliar os resultados obtidos nas tarefas de reconhecimento de faces, a identificação e a verificação. Apesar dessas tarefas constituírem uma das últimas etapas de um sistema de reconhecimento facial, vários outros módulos foram desenvolvidos, dos quais suas respectivas saídas são a entrada para estas tarefas. Para a execução dos conjuntos de experimentos, foram utilizadas as imagens dos bancos supracitados. As saídas geradas por cada etapa desenvolvida na metodologia serão descritas nas próximas subseções.

4.2.1 Geração de Imagens Sintéticas

O módulo de geração das imagens sintéticas é composto de três etapas principais e se caracteriza como um pré-processamento para este trabalho, uma vez que o módulo

pode ser executado em modo *batch* e suas saídas são utilizadas na etapa de validação. Nesse módulo, dada uma imagem frontal (Figuras 4.3 e 4.4), a face é detectada e o registro é realizado para a localização dos pontos fiduciais da face, conforme Figuras 4.5 e 4.6. A partir dos pontos fiduciais detectados, o 3DMM (Modelo Deformável 3D) é ajustado à face e 19 imagens sintéticas são geradas em diferentes poses, conforme Figuras 4.7 e 4.8. As imagens geradas são, então, utilizadas no módulo de validação.



Figura 4.3: Exemplo de imagem como parâmetro de entrada para geração de imagens sintéticas.



Figura 4.4: Exemplo de imagem como parâmetro de entrada para geração de imagens sintéticas.

Embora o primeiro módulo desempenhe em modo *batch*, no qual o tempo de processamento não é crítico, os resultados do módulo de validação são inerentes às saídas

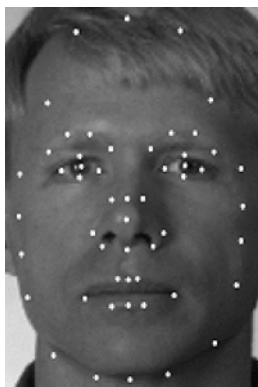


Figura 4.5: Exemplo da localização dos pontos fiduciais da Figura 4.3, recebida como parâmetro de entrada para geração de imagens sintéticas.

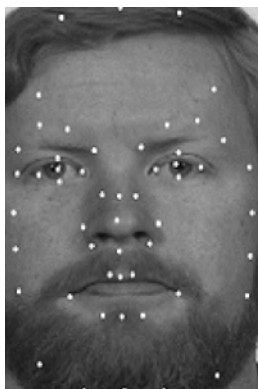


Figura 4.6: Exemplo a localização dos pontos fiduciais da Figura 4.4, recebida como parâmetro de entrada para geração de imagens sintéticas.

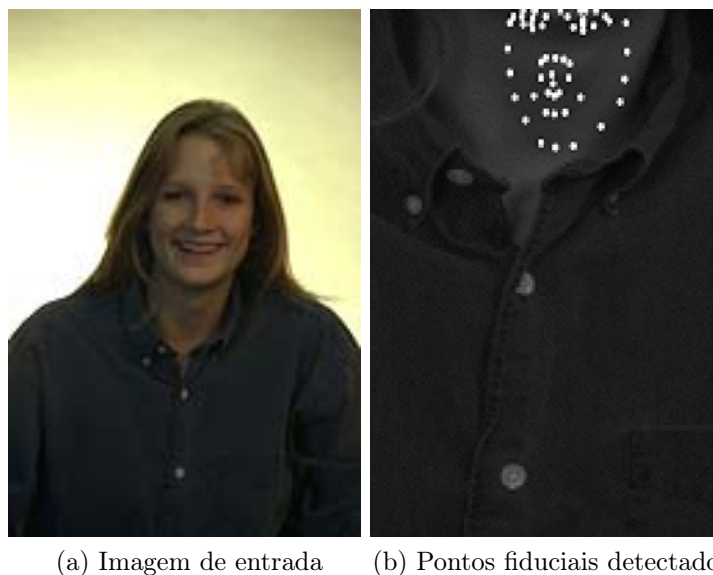


Figura 4.7: Exemplo das imagens sintéticas geradas a partir da Figura 4.3, recebida como parâmetro de entrada para o primeiro módulo.



Figura 4.8: Exemplo das imagens sintéticas geradas a partir da Figura 4.4, recebida como parâmetro de entrada para o primeiro módulo.

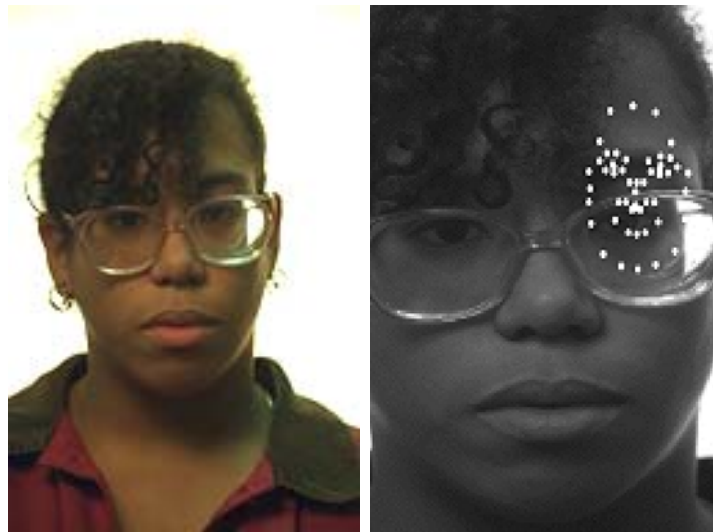
desse módulo. Dessa forma, para se obter resultados com melhor acurácia, a saída de cada etapa do primeiro módulo deve ter uma maior eficácia. Ainda que os algoritmos implementados em cada etapa do primeiro módulo tenham sido os que possuem maior acurácia, para alguns casos os resultados não foram adequados. As Figuras 4.9, 4.10 e 4.11 exemplificam alguns erros do algoritmo de detecção dos pontos fiduciais. Tais erros foram propagados para as próximas etapas, o que conseqüentemente gerou de forma errônea as imagens sintéticas, conforme exemplificado na Figura 4.12.



(a) Imagem de entrada (b) Pontos fiduciais detectados

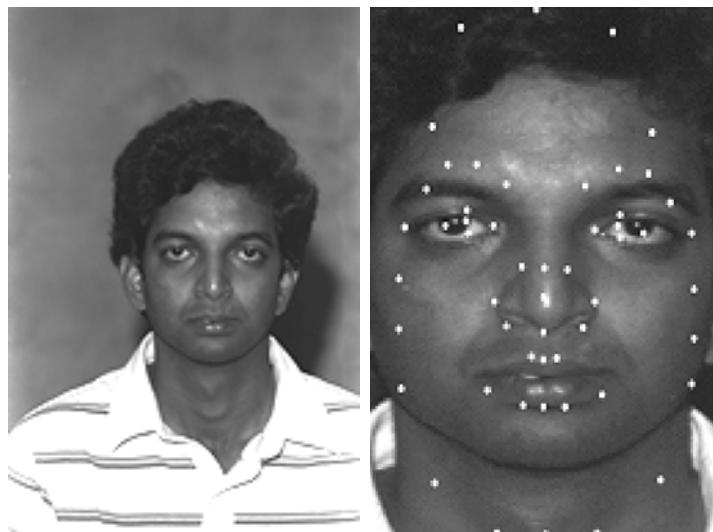
Figura 4.9: Exemplos de falha na etapa de detecção dos pontos fiduciais.

A partir do primeiro módulo, foram geradas as imagens sintéticas da partição *fa* do banco FERET compostas por faces frontais e das 40 imagens frontais do banco LATIM, que serão utilizadas nos experimentos.



(a) Imagem de entrada (b) Pontos fiduciais detectados

Figura 4.10: Exemplos de falha na etapa de detecção dos pontos fiduciais.



(a) Imagem de entrada (b) Pontos fiduciais detectados

Figura 4.11: Exemplos de falhas na etapa de detecção dos pontos fiduciais, nos quais os pontos ultrapassaram o contorno da face.

4.2.2 Reconhecimento Facial

A última etapa deste trabalho se caracteriza pela validação da abordagem proposta. Nesta etapa, foi realizado o reconhecimento de faces em poses arbitrárias por meio dos descritores locais. Dessa forma, a partir de uma imagem de teste, o ângulo da face é estimado de modo que o *matching* é realizado apenas com as imagens sintéticas geradas



Figura 4.12: Exemplo do erro propagado na etapa de detecção dos pontos fiduciais, no qual as imagens sintéticas foram geradas de forma errônea pelo fato do detector não ter localizado precisamente os pontos do contorno da face.

no mesmo ângulo. O algoritmo proposto por Zhu & Ramanan [2012] foi implementado neste trabalho para a inferência do ângulo, realizando também a detecção dos pontos fiduciais utilizados para a extração dos descritores. Sua precisão na estimativa do ângulo possui um desvio de $\pm 15^\circ$ e sua inferência está em função da face detectada. Em avaliação empírica, foi verificado que 5% das imagens submetidas para estimar a pose não tiveram a face detectada, e conseqüentemente não foi possível estimar o ângulo. No entanto, esse resultado está em função das imagens sintéticas geradas, o que pode ter sido influenciado pelos algoritmos das etapas do módulo anterior. Em adição à análise das saídas geradas pelo algoritmo, foi observado que os pontos fiduciais detectados possuem uma discretização em sua localização espacial devido à característica do algoritmo, conforme exemplificado na Figura 4.13. A metodologia desse algoritmo é fundamentada em modelos nos quais os pontos fiduciais possuem suas localizações fixas. Dessas localizações, algumas estão fora da região facial, o que influencia na descrição da face por meio destes pontos, uma vez que os descritores são extraídos da região do ponto localizado.

4.3 Experimentos

Os experimentos com as imagens dos bancos supracitados foram divididos em cinco categorias para validação da abordagem proposta:

4.3.1 Variação do tamanho do bloco

Essa categoria de testes busca validar o quão significativo é o tamanho do bloco utilizado para a extração dos descritores. Nesse cenário de teste, foi validada a taxa de verificação com descritores extraídos de blocos com tamanhos diferentes. Os tamanhos

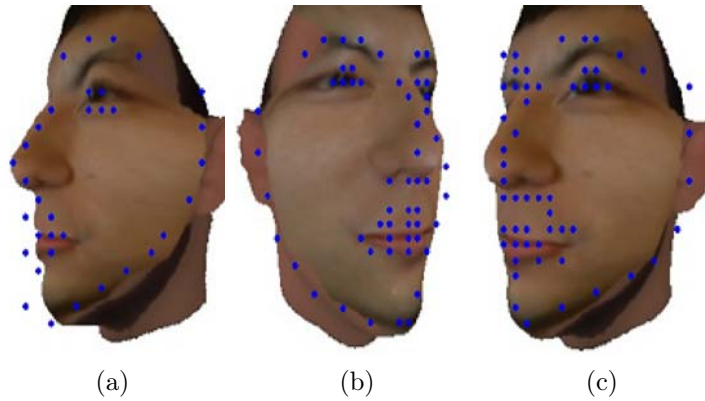


Figura 4.13: Exemplos dos pontos fiduciais detectados para a extração dos descritores, no qual a localização espacial dos pontos é discreta.

em pixels dos blocos foram definidos em 10×10 , 15×15 e 20×20 e as 761 imagens do conjunto *qr* foram utilizadas para a validação.

O gráfico apresentado na Figura 4.14 mostra a curva ROC (*Receiver Operation Characteristic*) dos resultados obtidos neste teste. A curva é discriminada pela taxa de Falsa Aceitação versus taxa de Correta Aceitação, ou seja, a relação entre especificidade e sensibilidade. Pelos resultados apresentados na curva, foi observado que a taxa de reconhecimento não teve alteração significativa entre o bloco de tamanho 15×15 e o bloco de tamanho 20×20 , do qual se obteve melhores resultados. De acordo com a Figura 4.14, é interessante destacar que nem sempre uma maior região resulta num melhor resultado, ou seja, o aumento no tamanho da região tende a não alterar o resultado final obtido para um certo tamanho de região, o que é indício de que as abordagens locais são melhores que as abordagens holísticas, uma vez que há informações desnecessárias em certas regiões da face. Além disto, a Área sob a Curva ROC (AUC) foi calculada numericamente pela regra do trapézio para os resultados de cada curva plotada. O bloco de tamanho 20×20 teve uma AUC de 0.72, enquanto que os blocos de tamanho 15×15 e 10×10 possuem uma AUC de 0.70 e 0.57, respectivamente. Pelo fato dos melhores resultados terem sido obtidos com um bloco de tamanho 20×20 , para os demais experimentos realizados neste trabalho foi mantido esse tamanho.

4.3.2 Reconhecimento por regiões

Embora a quantidade de regiões relacionadas aos pontos fiduciais implica na eficácia do sistema, algumas dessas regiões não são significativas e até mesmo podem gerar descritores que degradam a acurácia da abordagem proposta. Dessa forma, essa categoria de

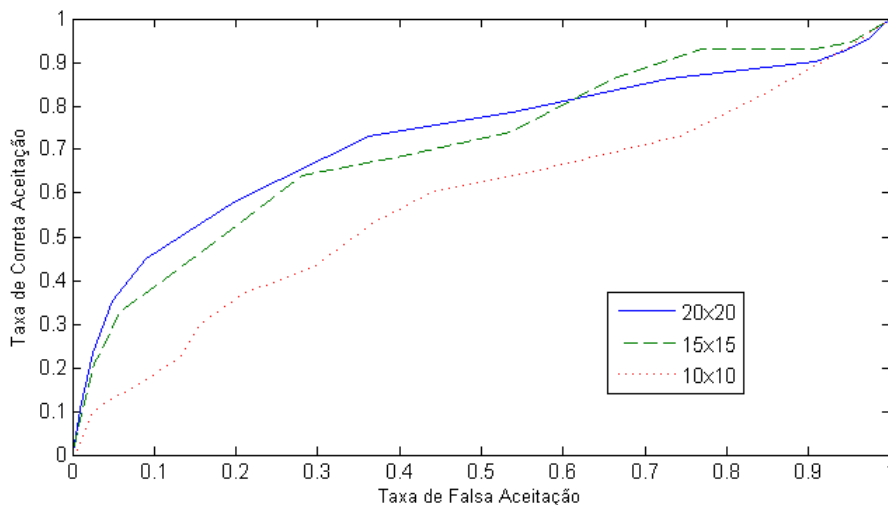


Figura 4.14: ROC da taxa de verificação do conjunto de imagens *qr* do banco FERET para diferentes tamanhos de blocos utilizados no descritor.

teste tem por finalidade a validação da representatividade de cada uma das 68 regiões no processo de reconhecimento de faces. Para realizar essa validação, foi utilizada a partição *fb* do banco FERET, sendo esta composta por 992 imagens de teste.

A Figura 4.15 ilustra os resultados obtidos do classificador para cada região extraída em torno dos 68 pontos fiduciais. Mais especificamente, no eixo das abscissas, encontra-se a identificação de cada ponto fiducial, enquanto que no eixo das ordenadas, encontra-se o resultado da similaridade retornada pelo classificador. Para cada ponto, foi realizada a média dos resultados retornados pelo classificador, quando a uma região de uma imagem é incorretamente classificada (Media Não GroundTruth) e para as regiões das imagens corretamente classificadas (Media GroundTruth). A partir dos resultados obtidos neste experimento, nota-se que algumas regiões da face não são discriminativas, o que pode degradar a classificação final. Em observação manual dessas regiões, foi verificado que elas estão localizadas no contorno facial, ou seja, informações do plano de fundo acabam sendo inclusas na extração dos blocos, uma vez que o bloco é extraído a partir do ponto fiducial localizado.

4.3.3 Validação de Grandes Rotações

Imagens de face com grandes rotações são comumente utilizadas para validação das abordagens de reconhecimento de faces em múltiplas poses. Apesar disso, não há nenhuma definição para distinção entre pequenas e grandes rotações. Segundo Niinuma et al. [2013], as grandes rotações podem ser caracterizadas por imagens em que os dois

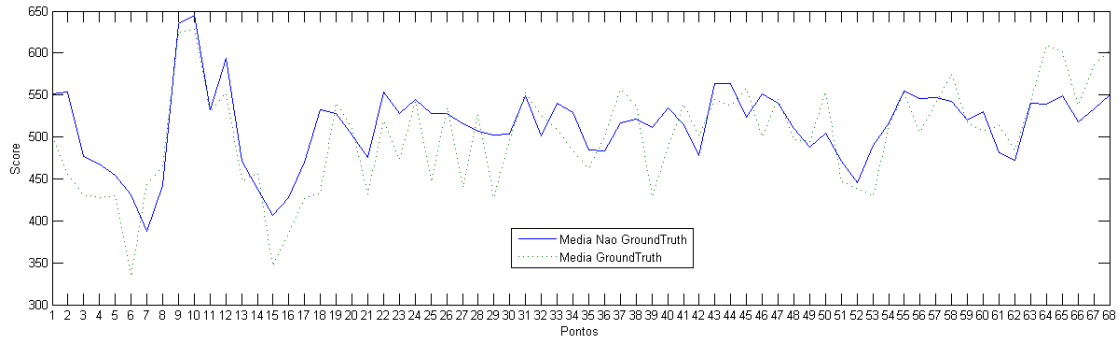


Figura 4.15: Avaliação dos resultados em cada região para o conjunto de imagens *fb* do banco FERET.

olhos da face não estão visíveis na imagem da face. A partir dessa consideração, neste cenário foram utilizadas 1320 imagens da partição *hr* do banco FERET com faces rotacionadas em 67.5° para a classificação. Além da validação do algoritmo proposto em grandes rotações, neste experimento também foi realizada a comparação da taxa de acerto com a abordagem holística, do qual se extraiu o histograma de toda a imagem para realizar a classificação.

O gráfico apresentado na Figura 4.16 mostra as curvas ROC resultantes dos experimentos deste teste, e a comparação dos resultados entre a abordagem local utilizada neste trabalho versus uma abordagem holística.

Neste gráfico, foi observado que as taxas de reconhecimento diminuíram de forma significativa para imagens com grandes rotações comparado ao gráfico da figura 4.14 que representa as pequenas rotações. Essa queda nas taxas se deve às poucas informações que discriminam a face de perfil e, sobretudo, pelo fato das imagens sintéticas terem sido geradas a partir de uma imagem frontal, a qual não possui todas as informações do perfil facial. Além disto, em comparação com a abordagem local, a qual foi utilizada neste trabalho, a abordagem holística não obteve resultados melhores, o que justifica que a utilização de abordagens locais para o reconhecimento de faces pode resultar em taxas melhores.

4.3.4 Validação das Abordagens Holísticas

Esta categoria de testes tem por objetivo a comparação das taxas das diferentes abordagens holísticas utilizadas na etapa de reconhecimento das faces. Dentre essas, foram desenvolvidas as abordagens Análise de Componentes Principais (PCA), Análise de Discriminante Linear (LDA), Análise de Componentes Principais em duas dimensões (2DPCA) e Histograma do Padrão Binário Local (LBPH). Nesta categoria, foram utili-

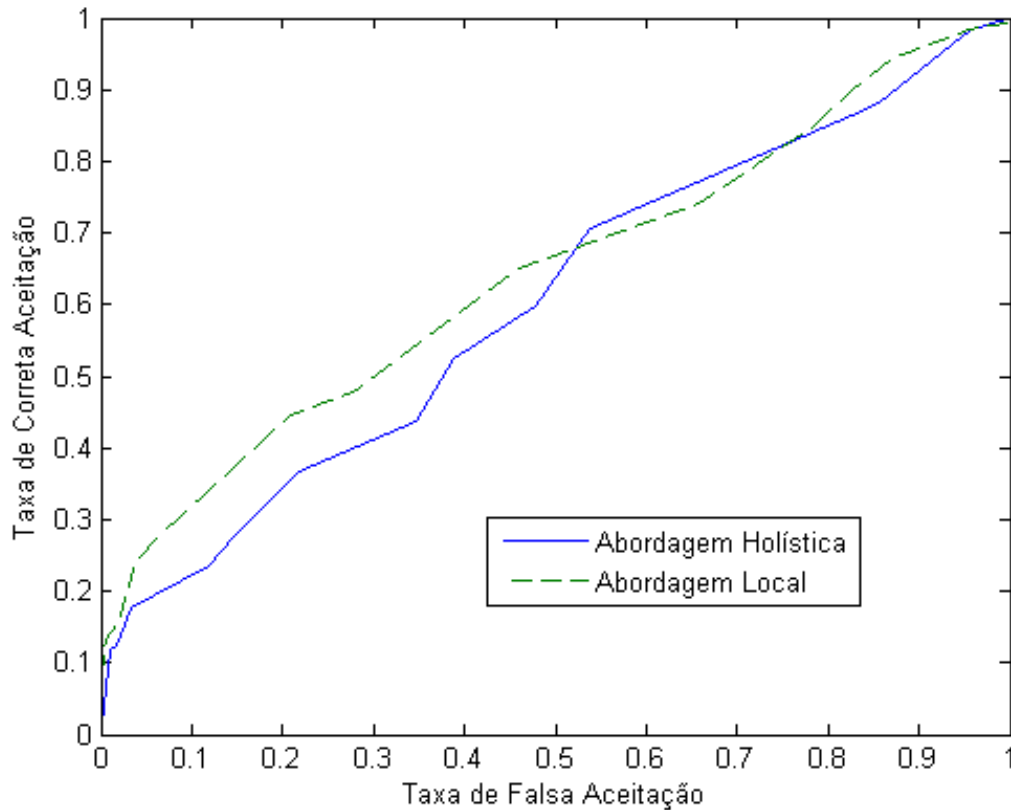


Figura 4.16: Comparação dos resultados para abordagem holística versus local em imagens da partição *hr* do banco FERET, categorizadas como grandes rotações.

zados o conjunto de imagens do banco LATIM composto por rotações de 30° , 45° e 90° para o teste e todas as imagens sintéticas geradas para o treino. A tabela 4.1 contém os resultados obtidos em cada abordagem e a partir dos resultados é possível verificar o impacto que as rotações têm sobre as taxas de reconhecimento nas abordagens holísticas.

Abordagem	Taxa de Reconhecimento
PCA	1.66%
LDA	1.66%
LBPH	3.33%
2DPCA	4.16%

Tabela 4.1: Resultados obtidos para a tarefa de verificação de faces nas diferentes abordagens.

4.3.5 Comparação com as Abordagens Holísticas

Na última categoria dos experimentos, foram avaliadas as abordagens tradicionais da literatura, PCA, LDA e 2DPCA, das quais as taxas de reconhecimento de faces foram comparadas com a da abordagem proposta. Para esse experimento, foram executados três conjuntos de testes de modo a avaliar as diferentes taxas à medida que a face é rotacionada. Foram utilizados o conjunto de imagens do banco LATIM com faces frontais e os conjuntos de faces rotacionadas à 30° e 45° para os testes. Os gráficos das Figuras 4.17, 4.18 e 4.19 mostram a curva ROC de cada abordagem para os diferentes ângulos.

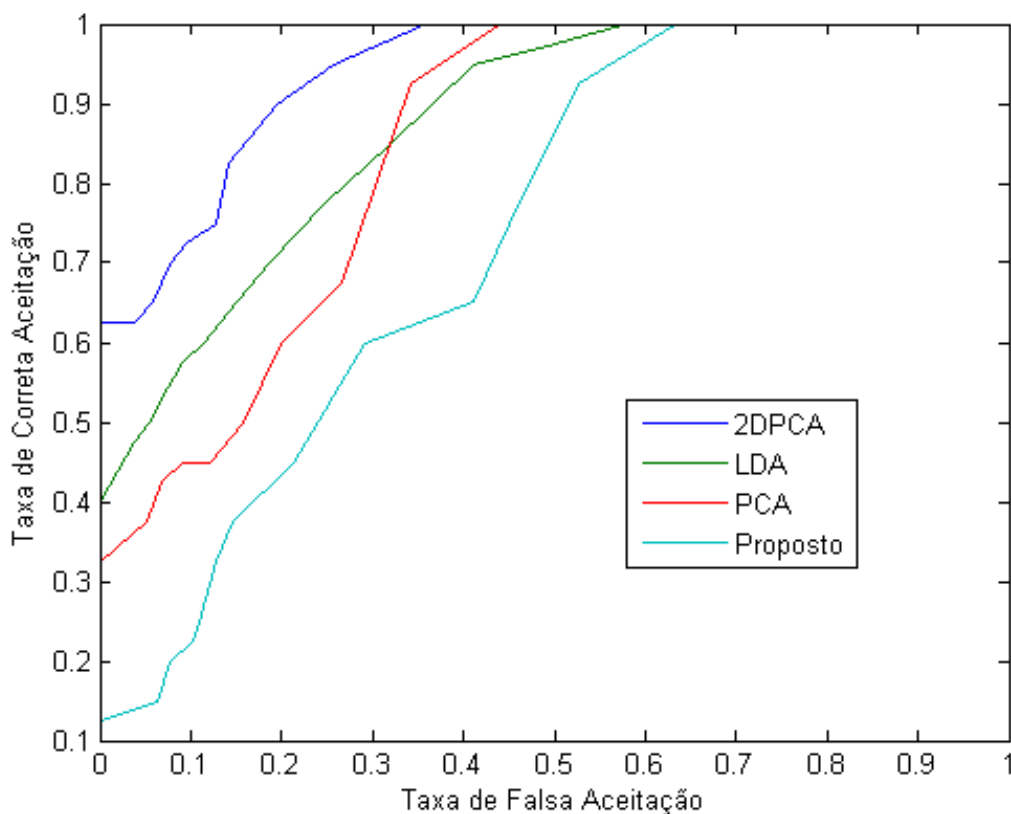


Figura 4.17: Comparação dos resultados para as abordagens tradicionais versus a abordagem proposta em faces frontais do banco LATIM.

A partir dos resultados apresentados nas curvas dos gráficos supracitados, foi verificado que a abordagem proposta alcançou taxas de reconhecimento superiores aos resultados das tradicionais abordagens de reconhecimento de faces à medida que o ângulo de rotação da face aumenta. Essa afirmação é observada por meio do gráfico da Figura 4.20, na qual foram extraídas as taxas de reconhecimento das técnicas em 0.1,

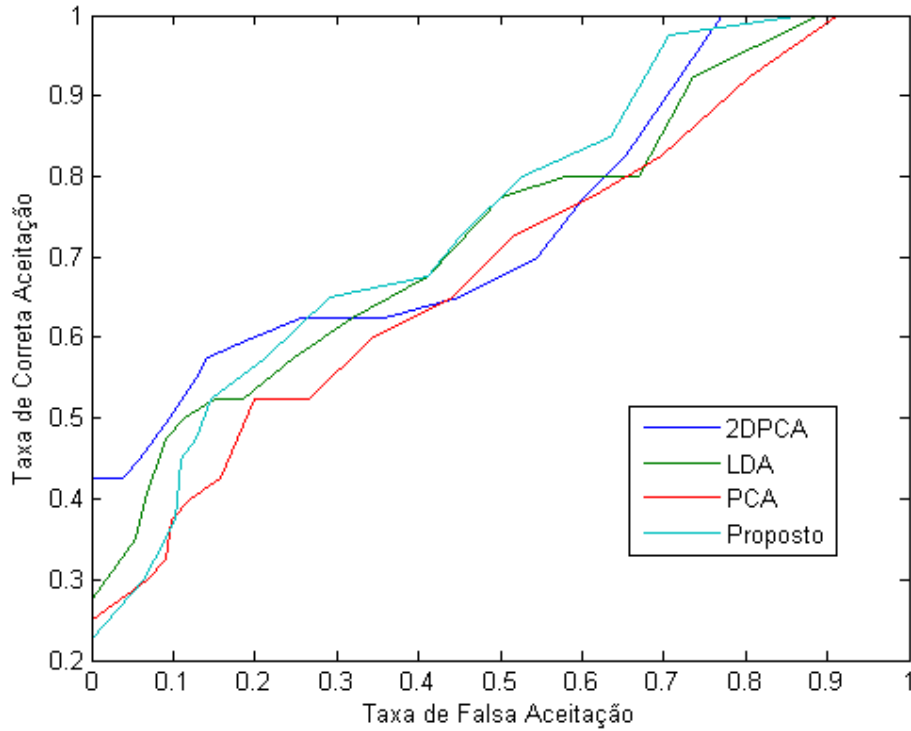


Figura 4.18: Comparação dos resultados para as abordagens tradicionais versus a abordagem proposta em faces rotacionadas à 30° do banco LATIM.

ou seja, o reconhecimento à uma margem de 10% de erro. Além disto, por meio do cálculo da AUC em cada curva plotada, a abordagem 2DPCA foi predominantemente melhor do que as outras abordagens para os testes em imagens com pequenas rotações, enquanto que o valor da AUC para o método proposto foi superior aos valores das abordagens da literatura quando a face possui uma rotação superior do que 30° , conforme a Tabela 4.2.

	Teste 1	Teste 2	Teste 3
2DPCA	0.94	0.74	0.60
LDA	0.87	0.62	0.52
PCA	0.84	0.59	0.53
Proposto	0.73	0.69	0.65

Tabela 4.2: Área sob a curva (AUC) para cada abordagem de cada experimento.

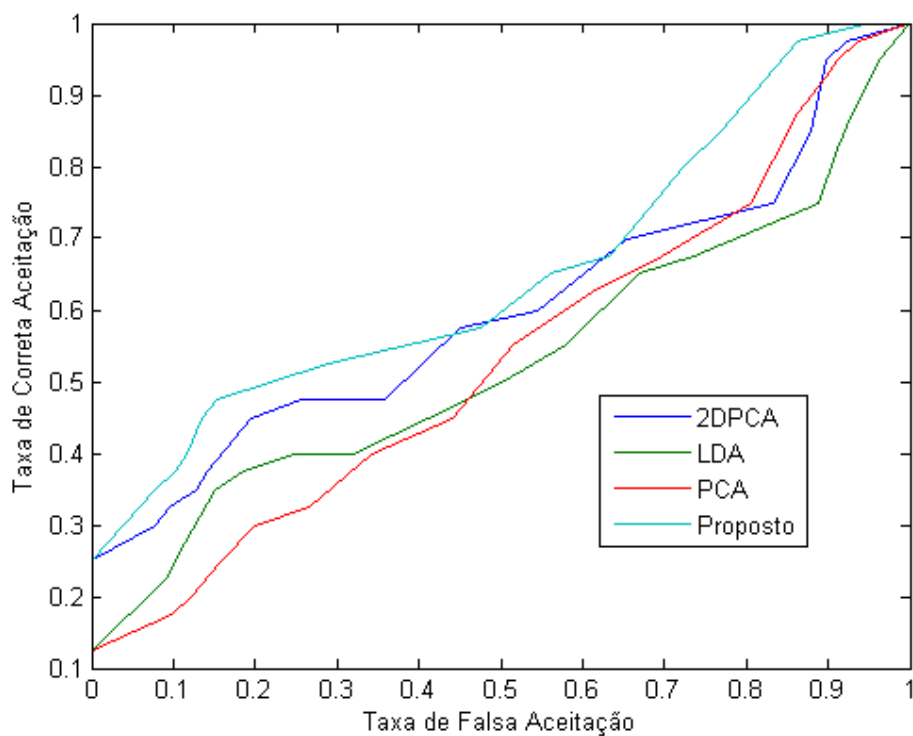


Figura 4.19: Comparação dos resultados para as abordagens tradicionais versus a abordagem proposta em faces rotacionadas à 45° do banco LATIM.

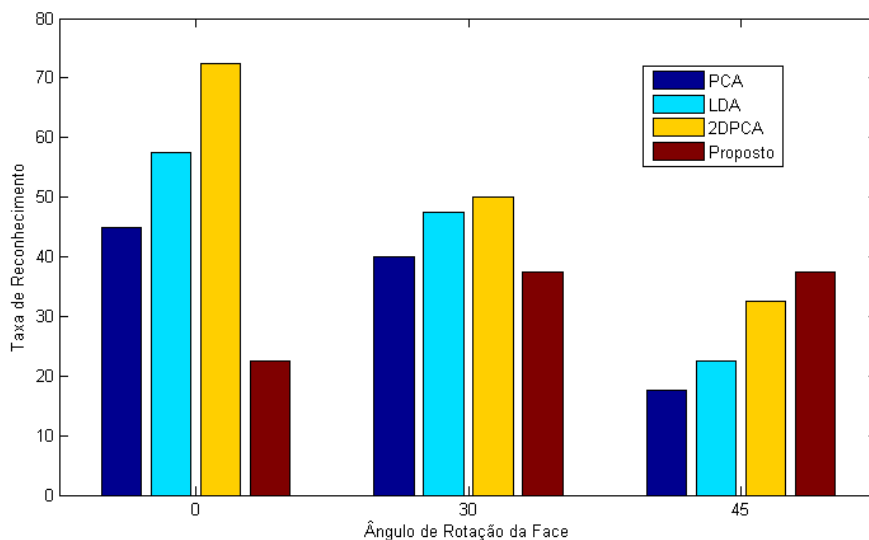


Figura 4.20: Comparação dos resultados para o reconhecimento de faces em diferentes ângulos à uma taxa de 10% de erro.

Capítulo 5

Conclusão

O Reconhecimento de Faces de forma automática ainda é um desafio para os atuais sistemas biométricos. Esse desafio se dá por fatores que abrangem desde o processo de aquisição das faces até o seu processamento. Um desses fatores é a rotação da face que apresenta maior degradação na acurácia dos sistemas devido às restrições impostas pela tecnologia. Uma das restrições refere-se à dificuldade em se ter várias imagens em diferentes ângulos de um mesmo indivíduo.

Neste trabalho, foi proposta uma abordagem para o Reconhecimento de Faces em poses arbitrárias. Essa abordagem é baseada em descritores locais extraídos dos pontos fiduciais localizados na face. Além disso, são realizados os dois principais processos de classificação de padrões: a verificação e a identificação. Para atingir os objetivos, foi desenvolvido um sistema para reconhecimento automático de faces composto por dois módulos. O primeiro módulo é responsável por eliminar a limitação de se ter apenas uma imagem de um indivíduo. Nesse módulo, uma imagem frontal é recebida como entrada e imagens sintéticas da face são geradas em múltiplas poses. Para a geração das imagens sintéticas, o sistema necessita da detecção dos pontos fiduciais nos quais o algoritmo ASM de código-livre foi adaptado à abordagem proposta e os pontos fiduciais detectados foram utilizados no algoritmo de ajuste da face a um modelo 3D canônico. Este algoritmo foi desenvolvido de acordo com a abordagem proposta por Park et al. [2010], porém a função de minimização foi resolvida numericamente através dos sistemas não-lineares. O segundo módulo, por sua vez, consistiu no processo de validação da abordagem proposta na qual são realizadas as duas tarefas de reconhecimento facial, a identificação e a verificação. Esse módulo basicamente recebe uma imagem de teste como entrada na qual a face está rotacionada num ângulo qualquer, seu ângulo é inferido e os pontos fiduciais são detectados. A partir do ângulo extraído, a última etapa deste módulo considera apenas as imagens sintéticas geradas pelo módulo anterior que

tenham o mesmo ângulo estimado, e descritores locais são extraídos da imagem de teste para o *matching* com os descritores extraídos previamente das imagens sintéticas.

A abordagem proposta foi avaliada em diferentes cenários a partir dos quais foi verificado que os resultados deste trabalho são promissores tanto para pequenas rotações quanto para grandes rotações, de acordo com os valores das AUC gerados pelos experimentos executados. Além disto, foi verificado que as informações faciais em alguns pontos da face não são significativas para realizar a descrição de um indivíduo. Dessa forma, as abordagens locais são substancialmente melhores para os problemas de reconhecimento facial em múltiplas poses. Apesar dos resultados não terem alcançado as taxas de reconhecimento da literatura, algumas contribuições pontuais foram efetuadas. Dessas, destaca-se a análise do quão necessária é a quantidade de informação fornecida para os algoritmos de classificação e sobretudo, a importância em se ter uma eficiente abordagem de detecção para uso nos algoritmos de reconhecimento de faces.

O reconhecimento de faces em múltiplas poses é um problema complexo e ainda em aberto para os atuais sistemas de reconhecimento. No desenvolvimento desta dissertação, alguns pontos não puderam ser avaliados devido ao volume do trabalho desenvolvido, ficando como tarefas para os trabalhos futuros. Dentre esses, podemos destacar a utilização de outras características faciais como descritores para o processo de *matching*. As informações de contorno do perfil facial, segundo Zhang & Gao [2009], são altamente descritivas principalmente a energia de curvatura do contorno do perfil facial. Uma medida de similaridade foi utilizada para realizar a classificação entre os descritores extraídos. A partir disto, outra tarefa considerada como trabalho futuro é a avaliação de outros classificadores, como por exemplo SVM e outras medidas de similaridade. O modelo 3D sintético de uma face qualquer foi gerado por meio de uma imagem com a face frontal. Dado que uma face frontal não possui todas as informações do perfil facial, o modelo quando rotacionado, não trará uma projeção com a real textura da face. Desta forma, a utilização das informações de imagens com a face em perfil para sintetização do modelo seria uma outra tarefa futura a ser desenvolvida, o que poderá trazer melhores resultados para a classificação. Por fim, outro trabalho futuro seria avaliar se a localização dos pontos fiduciais de forma manual têm influência direta no resultados, uma vez que a acurácia dos atuais detectores não é confiável.

Referências Bibliográficas

- Abate, A. F.; Nappi, M.; Riccio, D. & Sabatino, G. (2007). 2d and 3d face recognition: A survey. *Pattern Recognition Letters*, 28(14):1885 – 1906. Image: Information and Control.
- Ahonen, T.; Hadid, A. & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041.
- Araujo, B. & Machado, A. (2011). Towards a better comprehension of the role of image registration in face recognition algorithms. Em *Biometric Measurements and Systems for Security and Medical Applications (BIOMS), 2011 IEEE Workshop on*, pp. 1 –6.
- Ba, S. & Odobez, J. (2004). A probabilistic framework for joint head tracking and pose estimation. Em *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pp. 264 – 267 Vol.4.
- Blanz, V. & Vetter, T. (1999). A morphable model for the synthesis of 3d faces. Em *Proceedings of the 26th annual conference on Computer graphics and interactive techniques, SIGGRAPH '99*, pp. 187--194, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.
- Blanz, V. & Vetter, T. (2003). Face recognition based on fitting a 3d morphable model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1063–1074.
- Breitenstein, M.; Kuettel, D.; Weise, T.; van Gool, L. & Pfister, H. (2008). Real-time face pose estimation from single range images. Em *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1 –8.
- Brunelli, R. & Poggio, T. (1993). Face recognition: features versus templates. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(10):1042–1052.

- Cai, Q.; Sankaranarayanan, A.; Zhang, Q.; Zhang, Z. & Liu, Z. (2010). Real time head pose tracking from multiple cameras with a generic model. Em *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pp. 25 –32.
- Demir, E.; Akarun, L. & Alpaydin, E. (2000). Two-stage approach for pose invariant face recognition. Em *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on*, volume 6, pp. 2342–2344 vol.4.
- Fanelli, G.; Gall, J. & Van Gool, L. (2011). Real time head pose estimation with random regression forests. Em *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 617 –624.
- Fanelli, G.; Gall, J. & Van Gool, L. (2012). Real time 3d head pose estimation: Recent achievements and future challenges. Em *Communications Control and Signal Processing (ISCCSP), 2012 5th International Symposium on*, pp. 1 –4.
- Fischer, M.; Ekenel, H. & Stiefelhagen, R. (2012). Analysis of partial least squares for pose-invariant face recognition. Em *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, pp. 331–338.
- Gao, Y. & Leung, M. K. H. (2002). Face recognition using line edge map. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(6):764–779.
- Gee, A. & Cipolla, R. (1996). Fast visual tracking by temporal consensus. *Image and Vision Computing*, 14(2):105 – 114.
- Gong, S.; McKenna, S. & Psarrou, A. (2000). *Dynamic Vision: From Images to Face Recognition*. Image Processing. Imperial College Press.
- Guo, H.; Schwartz, W. R. & Davis, L. S. (2011). Face Verification using Large Feature Sets and One Shot Similarity. Em *International Joint Conference on Biometrics*.
- He, G.-H.; Zhu, B. & ying Gan, J. (2008). 3-d face reconstruction using rbf and b spline method. *Image and Signal Processing, Congress on*, 1:239–243.
- Huang, C.; Ding, X. & Fang, C. (2010). Head pose estimation based on random forests for multiclass classification. Em *Pattern Recognition (ICPR), 2010 20th International Conference on*, pp. 934 –937.
- Huang, D.; Storer, M.; De la Torre, F. & Bischof, H. (2011). Supervised local subspace learning for continuous head pose estimation. Em *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 2921 –2928.

- Jain, A.; Klare, B. & Park, U. (2011). Face recognition: Some challenges in forensics. Em *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pp. 726 –733.
- Lanitis, A. (2004). Person identification from heavily occluded face images. Em *Proceedings of the 2004 ACM symposium on Applied computing, SAC '04*, pp. 5--9, New York, NY, USA. ACM.
- Levine, M. D. & Yu, Y. C. (2009). State-of-the-art of 3d facial reconstruction methods for face recognition based on a single 2d training image per person. *Pattern Recognition Letters*, 30(10):908 – 913.
- Li, A.; Shan, S. & Gao, W. (2012). Coupled bias - variance tradeoff for cross-pose face recognition. *Image Processing, IEEE Transactions on*, 21(1):305–315.
- Li, S. Z. & Jain, A. K. (2004). *Handbook of Face Recognition*. Springer.
- Liang, X. & Tong, W. (2012). Face pose estimation using near-infrared images. Em *Communication Systems and Network Technologies (CSNT), 2012 International Conference on*, pp. 216 –220.
- Malassiotis, S. & Srinivasan, M. G. (2005). Robust real-time 3d head pose estimation from range data. *Pattern Recognition*, 38:1153--1165.
- Martins, P. A. D. (2008). Active appearance models for facial expression recognition and monocular head pose estimation.
- Mckenna, S.; Gong, S. & Raja, Y. (1997). Face recognition in dynamic scenes. Em *Proceedings of British Machine Vision Conference*, pp. 140--151. BMVA Press.
- Milborrow, S. & Nicolls, F. (2008). Locating facial features with an extended active shape model. Em *Proceedings of the 10th European Conference on Computer Vision: Part IV, ECCV '08*, pp. 504--513, Berlin, Heidelberg. Springer-Verlag.
- Ming, Y.; Ruan, Q. & Li, X. (2010). 3d face reconstruction using a single 2d face image. Em *Educational and Information Technology (ICEIT), 2010 International Conference on*, volume 3, pp. V3-32 –V3-36.
- Murphy-Chutorian, E. & Trivedi, M. (2009). Head pose estimation in computer vision: A survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(4):607 –626.

- Niinuma, K.; Han, H. & Jain, A. (2013). Automatic multi-view face recognition via 3d model based pose regularization. Em *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pp. 1–8.
- Ojala, T.; Pietikainen, M. & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *TPAMI*, 24(7):971–987.
- Ozturk, O.; Yamasaki, T. & Aizawa, K. (2009). Tracking of humans and estimation of body/head orientation from top-view single camera for visual focus of attention analysis. Em *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pp. 1020–1027.
- Park, U.; Tong, Y. & Jain, A. (2010). Age-invariant face recognition. *TPAMI*, 32(5):947–954.
- Pentland, A.; Moghaddam, B. & Starner, T. (1994). View-based and modular eigenspaces for face recognition. Em *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, pp. 84–91.
- Phillips, P.; Wechsler, H.; Huang, J. & Rauss, P. J. (1998). The {FERET} database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295 – 306.
- Pietikäinen, M.; Zhao, G.; Hadid, A. & Ahonen, T. (2011). *Computer Vision Using Local Binary Patterns*. Number 40 in Computational Imaging and Vision. Springer.
- Prince, S. J. D.; Elder, J. H.; Warrell, J. & Felisberti, F. M. (2008). Tied factor analysis for face recognition across large pose differences. *TPAMI*, 30(6):970–984.
- Rydfalk, M. (1987). Candide, a parameterized face. Relatório técnico LiTH-ISY-I-866, Dept. of Electrical Engineering, Linköping University, Sweden.
- Shahdi, S. & Abu-Bakar, S. A. R. (2012). Varying pose face recognition using combination of discrete cosine amp; wavelet transforms. Em *Intelligent and Advanced Systems (ICIAS), 2012 4th International Conference on*, volume 2, pp. 642–647.
- Sharma, A. & Jacobs, D. (2011). Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch. Em *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 593–600.

- Simon, D.; Hebert, M. & Kanade, T. (1994). Real-time 3-d pose estimation using a high-speed range sensor. Em *Robotics and Automation, 1994. Proceedings., 1994 IEEE International Conference on*, pp. 2235 –2241 vol.3.
- Siriteerakul, T.; Sato, Y. & Boonjing, V. (2011). Estimating change in head pose from low resolution video using lbp-based tracking. Em *Intelligent Signal Processing and Communications Systems (ISPACS), 2011 International Symposium on*, pp. 1 –6.
- Suen, C.; Langaroudi, A.; Feng, C. & Mao, Y. (2007). A survey of techniques for face reconstruction. Em *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, pp. 3554 –3560.
- Tan, X.; Chen, S. & hua Zhou Fuyan Zhang, Z. (2006). Face recognition from a single image per person: A survey. *Pattern Recognition*, 39:1725--1745.
- Tang, Y.; Sun, Z. & Tan, T. (2011). Face pose estimation based on integral slice features of single depth images. Em *Pattern Recognition (ACPR), 2011 First Asian Conference on*, pp. 530 –534.
- Turk, M. & Pentland, A. (1991a). Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1):71--86.
- Turk, M. & Pentland, A. (1991b). Face recognition using eigenfaces. Em *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, pp. 586 –591.
- Yu, Y. (2007). A method for synthesis of three dimensional head with single 2d image. Dissertação de mestrado, McGill University, Canada.
- Zhang, X. & Gao, Y. (2009). Face recognition across pose: A review. *Pattern Recognition*, 42(11):2876 – 2896.
- Zhao, W.; Rosenfeld, R. C. A. & Phillips, P. (2003). Face recognition: A literature survey. *ACM Computer*, 35(4):399--458.
- Zhu, X. & Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild. Em *Computer Vision and Pattern Recognition (CVPR)*, Providence, Rhode Island.

Apêndice A

Algoritmo Ajuste da Face

O ajuste do modelo 3D à uma imagem 2D é baseado na presunção de que é possível reconstruir uma face 3D por meio do modelo 3DMM e seus respectivos coeficientes $\vec{\alpha}$ de controle da forma. O modelo 3DMM simplificado é representado pela seguinte equação:

$$\vec{S}_\alpha = \vec{S}_{mm} + \sum_{l=1}^L \alpha_l W_l. \quad (\text{A.1})$$

Esse ajuste é realizado utilizando os pontos fiduciais detectados na imagem 2D que são correspondentes aos pontos(vértices) do modelo simplificado. Destarte, a seguinte função de minimização é calculada com o intuito de encontrar o conjunto de transformações (rotação, translação e escala) que a face 2D, representada pelos pontos, sofreu durante o processo de aquisição da imagem:

$$E(P, R, t, a, \alpha_{l=1}^L) = \|S_{2d,i}^j - T_{P,R,t,a}(S_\alpha)\|^2. \quad (\text{A.2})$$

Como os pontos detectados na face 2D não são precisos em relação aos seus correspondentes no modelo, a função de minimização é controlada por um termo de regularização de *Tikhonov*:

$$E' = E/\sigma^2 + \sum \alpha_l^2/\lambda_{l=1}^L. \quad (\text{A.3})$$

O processo de minimização da equação A.3 é feito iterativamente onde se atualiza na primeira iteração os parâmetros de transformação geométrica e na próxima os parâmetros de controle da forma. O seguinte pseudo código exemplifica o método.

Algorithm 1 Ajuste do Modelo 3D à face 2D

Entrada: $S_{2d} = (x_1, y_1, x_2, \dots, y_n)^T$

Saída: $S_{3d} = (X_1, Y_1, Z_1, X_2, \dots, Y_n, Z_n)^T$

1. Inicialização:

$$\vec{\alpha} = 0, T = 0, R = I, k = 1$$

2. Alinhamento:

S_{2d} é alinhado com o modelo médio 2D obtido pela projeção ortográfica do modelo frontal 3D médio.

3. Atualiza R e T , com os parâmetros $\vec{\alpha}$ fixos.

$$\operatorname{argmin}(R, T)_k = \|P(RS + T) - S_{2d}\|_{\alpha=\alpha_{k-1}}^2$$

4. Atualiza os parâmetros $\vec{\alpha}$ com os parâmetros R e T fixos.

$$\operatorname{argmin}(\alpha_k) = \|P(RS + T) - S_{2d}\|_{(R,T)=(R,T)_k}^2$$

5. Reconstruir o modelo 3DMM com os parâmetros $\vec{\alpha}$ estimados

6. Se:

$$\|P(RS + T) - S_{2d}\|_{(R,T)=(R,T)_k} < \epsilon$$

Senão, $k = k + 1$, retorne ao passo 3.

7. Reconstruir a face 3D com os parâmetros $\vec{\alpha}$ estimados.

Apêndice B

Fórmulas de Minimização

A equação B.1

$$E = \left\| P(R(\vec{S} + \vec{T})) - \vec{S}_{2D} \right\|^2. \quad (\text{B.1})$$

onde $P = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix}$ representa a matriz de projeção ortográfica, $R = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}$ a

matriz de rotação com ângulo *yaw* 0, $\vec{S} = \begin{vmatrix} \vec{x} \\ \vec{y} \\ \vec{z} \end{vmatrix}$ é a matriz do modelo 3D, $\vec{T} = \begin{vmatrix} \vec{tx} \\ \vec{ty} \\ \vec{tz} \end{vmatrix}$

é uma matriz $3 \times n$ composta pelos n vetores $T = [t_x t_y t_z]^T$ de translação e \vec{S}_{2D} é o vetor $2 \times n$ transformado da matriz $2 \times n$ que representa os pontos fiduciais detectados na imagem.

A expansão da equação B.1 pode ser realizada da seguinte forma:

$$E = \left\| \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} \left(\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} \begin{vmatrix} \vec{x} \\ \vec{y} \\ \vec{z} \end{vmatrix} + \begin{vmatrix} \vec{tx} \\ \vec{ty} \\ \vec{tz} \end{vmatrix} \right) - \begin{vmatrix} \vec{Sx}_{2D} \\ \vec{Sy}_{2D} \end{vmatrix} \right\|^2$$

$$E = \left\| \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} \left(\begin{vmatrix} a\vec{x} & b\vec{y} & c\vec{z} \\ d\vec{x} & e\vec{y} & f\vec{z} \\ g\vec{x} & h\vec{y} & i\vec{z} \end{vmatrix} + \begin{vmatrix} \vec{tx} \\ \vec{ty} \\ \vec{tz} \end{vmatrix} \right) - \begin{vmatrix} \vec{Sx}_{2D} \\ \vec{Sy}_{2D} \end{vmatrix} \right\|^2$$

$$E = \left\| \begin{vmatrix} a\vec{x} & b\vec{y} & c\vec{z} \\ d\vec{x} & e\vec{y} & f\vec{z} \end{vmatrix} - \begin{vmatrix} \vec{Sx}_{2D} \\ \vec{Sy}_{2D} \end{vmatrix} \right\|^2$$

$$E = \sum_i^n (((a\vec{x} + b\vec{y} + c\vec{z} + tx)_i - Sx_{2Di})^2 + ((d\vec{x} + e\vec{y} + f\vec{z} + ty)_i - Sy_{2Di})^2) / \sigma^2 + \sum_{j=1}^m \alpha_j^2 / \lambda_j$$

Os vetores \vec{x} , \vec{y} e \vec{z} da equação E representam os vértices do modelo da equação

3.13 e são representados da seguinte forma:

$$\vec{x} = \vec{S}_x + \sum_{l=1}^m \alpha_l W_{l,x} \quad \vec{y} = \vec{S}_y + \sum_{l=1}^m \alpha_l W_{l,y} \quad \vec{z} = \vec{S}_z + \sum_{l=1}^m \alpha_l W_{l,z} \quad (\text{B.2})$$

A matriz de rotação R representa a composição das rotações nos eixos X (*pitch*) com ângulo θ , Y (*yaw*) com ângulo ρ e Z (*roll*) com ângulo ϕ , respectivamente, e é dada pela seguinte equação:

$$R = \begin{vmatrix} \cos\phi\cos\rho & (-\cos\phi\sen\rho) & \sen\phi \\ (\sen\theta\sen\phi\cos\rho + \cos\theta\sen\rho) & (-\sen\theta\sen\phi\sen\rho + \cos\theta\cos\rho) & (-\sen\theta\cos\phi) \\ (-\cos\theta\sen\phi\cos\rho + \sen\theta\sen\rho) & (\cos\theta\sen\phi\sen\rho + \sen\theta\cos\rho) & \cos\theta\cos\phi \end{vmatrix} \quad (\text{B.3})$$

Com base nas definições acima, a equação de minimização E é definida como:

$$\begin{aligned} E = \sum_i^n & \left((\cos\phi\cos\rho(\vec{S}_x + \sum_{l=1}^m \alpha_l W_{l,x}) - \cos\phi\sen\rho(\vec{S}_y + \sum_{l=1}^m \alpha_l W_{l,y}) + \sen\phi(\vec{S}_z + \sum_{l=1}^m \alpha_l W_{l,z}) + tx)_i - Sx_{2D_i} \right)^2 \\ & + \left((\sen\theta\sen\phi\cos\rho + \cos\theta\sen\rho)(\vec{S}_x + \sum_{l=1}^m \alpha_l W_{l,x}) + \right. \\ & \left. (-\sen\theta\sen\phi\sen\rho + \cos\theta\cos\rho)(\vec{S}_y + \sum_{l=1}^m \alpha_l W_{l,y}) + (-\sen\theta\cos\phi)(\vec{S}_z + \sum_{l=1}^m \alpha_l W_{l,z}) + ty \right)_i \\ & - Sy_{2D_i} \right)^2 / \sigma^2 + \sum_{j=1}^m \alpha_j^2 / \lambda_j \end{aligned} \quad (\text{B.4})$$

Calculando as derivadas parciais de E com respeito às variáveis θ , ϕ , ρ , tx , ty e α_j e igualando à zero, teremos um sistema não linear que pode ser solucionado numericamente.