UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

KELVIN SALTON DO PRADO

**Comparação de técnicas de reconhecimento facial para identificação de
presença em um ambiente real e semicontrolado**

São Paulo

2018

KELVIN SALTON DO PRADO

**Comparação de técnicas de reconhecimento facial para identificação de
presença em um ambiente real e semicontrolado**

Dissertação apresentada à Escola de Artes,
Ciências e Humanidades da Universidade de
São Paulo para obtenção do título de Mestre
em Ciências pelo Programa de Pós-graduação
em Sistemas de Informação.

Área de concentração: Metodologia e
Técnicas da Computação

Versão corrigida contendo as alterações
solicitadas pela comissão julgadora em 14
de novembro de 2017. A versão original
encontra-se em acervo reservado na Biblio-
teca da EACH-USP e na Biblioteca Digital
de Teses e Dissertações da USP (BDTD), de
acordo com a Resolução CoPGr 6018, de 13
de outubro de 2011.

Orientador: Prof. Dr. Norton Trevisan Roman

São Paulo

2018

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

CATALOGAÇÃO-NA-PUBLICAÇÃO

(Universidade de São Paulo. Escola de Artes, Ciências e Humanidades. Biblioteca)
CRB-8 4625

Prado, Kelvin Salton do

Comparação de técnicas de reconhecimento facial para
identificação de presença em um ambiente real e semicontrolado /
Kelvin Salton do Prado ; orientador, Norton Trevisan Roman. – 2017.
171 f. : il

Dissertação (Mestrado em Ciências) - Programa de Pós-
Graduação em Sistemas de Informação, Escola de Artes,
Ciências e Humanidades, Universidade de São Paulo.
Versão corrigida

1. Reconhecimento de padrões. 2. Inteligência artificial. 3.
Processamento de imagens. 4. Percepção da face. I. Roman,
Norton Trevisan, orient. II. Título.

CDD 22.ed.— 006.4

Dissertação de autoria de Kelvin Salton do Prado, sob o título “**Comparação de técnicas de reconhecimento facial para identificação de presença em um ambiente real e semicontrolado**”, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação, na área de concentração Metodologia e Técnicas da Computação, aprovada em 14 de novembro de 2017 pela comissão julgadora constituída pelos doutores:

Prof. Dr. Norton Trevisan Roman

Instituição: Escola de Artes, Ciências e Humanidades da Universidade de São Paulo
Presidente

Prof. Dr. Mauricio Marengoni

Instituição: Universidade Presbiteriana Mackenzie

Profa. Dra. Sarajane Marques Peres

Instituição: Escola de Artes, Ciências e Humanidades da Universidade de São Paulo

Prof. Dr. Marcelo de Souza Lauretto

Instituição: Escola de Artes, Ciências e Humanidades da Universidade de São Paulo

Eu dedico este trabalho aos meus pais, pois foi por meio do esforço e dedicação deles que eu consegui chegar até aqui.

Agradecimentos

Agradeço primeiramente aos meus familiares, principalmente aos meus pais Vanderlei Dornelles do Prado e Sandra Maria Salton do Prado por todo o apoio moral e financeiro durante o desenvolvimento deste trabalho, pois mesmo distante estavam sempre me dando o suporte necessário e me incentivando a seguir em frente. Agradeço a minha noiva Fhaira Meyer que me apoiou de todas as formas possíveis e me deu suporte em todos os momentos. Agradeço ao meu orientador Prof. Dr. Norton Trevisan Roman pelo auxílio, não só no desenvolvimento deste trabalho, mas também durante todo o meu percurso no curso de pós-graduação, inclusive me ensinando sobre o ofício de professor durante as tutorias em que participei. Agradeço também pela colaboração e o suporte da academia Central Kungfu-Wushu, em especial ao Prof. Enrique Ortega, pelos dados disponibilizados e pela prestatividade de sempre. Agradeço a Universidade de São Paulo, bem como aos professores do curso de pós-graduação em sistemas de informação, pela oportunidade de poder estudar em uma das melhores universidades do país e pela excelente infraestrutura disponibilizada aos seus alunos. Agradeço ao pessoal da empresa Condor Instruments pela compreensão e apoio durante o desenvolvimento deste trabalho. E finalmente agradeço a Deus pela oportunidade de conhecer todas estas pessoas e tornar possível o desenvolvimento deste trabalho.

“Nós só podemos ver um pouco do futuro, mas o suficiente para perceber que há muito a fazer.”

(Alan Mathison Turing)

Resumo

Prado, Kelvin Salton do. **Comparação de técnicas de reconhecimento facial para identificação de presença em um ambiente real e semicontrolado**. 2018. 171 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2018. Versão Corrigida.

O reconhecimento facial é uma tarefa que os seres humanos realizam naturalmente todos os dias e praticamente sem esforço nenhum. Porém para uma máquina este processo não é tão simples. Com o aumento do poder computacional das máquinas atuais criou-se um grande interesse no processamento de imagens e vídeos digitais, com aplicações nas mais diversas áreas de conhecimento. Este trabalho objetiva a comparação de técnicas de reconhecimento facial, já conhecidas na literatura, com o intuito de identificar qual técnica possui melhor desempenho em um ambiente real e semicontrolado. Secundariamente avalia-se a possibilidade da utilização de uma ou mais técnicas de reconhecimento facial para identificar automaticamente a presença de alunos em uma sala de aula de artes marciais, utilizando imagens das câmeras de vigilância instaladas no recinto, levando em consideração aspectos importantes, tais como: imagens com pouca nitidez, luminosidade não ideal, movimentação constante dos alunos e o fato das câmeras estarem em um ângulo fixo. Este trabalho está relacionado às áreas de Processamento de Imagens e Reconhecimento de Padrões, e integra a linha de pesquisa de “Monitoramento de Presença” do projeto “Ensino e Monitoramento de Atividades Físicas via Técnicas de Inteligência Artificial” (Processo 2014.1.923.86.4, publicado no DOE 125(45), em 10/03/2015), projeto este executado em conjunto da Universidade de São Paulo, Faculdade Campo Limpo Paulista e Academia Central Kungfu-Wushu. Com os experimentos realizados e apresentados neste trabalho foi possível concluir que, dentre os métodos de reconhecimento facial utilizados, o método *Local Binary Patterns* teve o melhor desempenho no ambiente proposto. Por outro lado, o método *Eigenfaces* teve o pior desempenho de acordo com os experimentos realizados. Além disso, foi possível concluir também que não é viável a realização da detecção de presença automática de forma confiável no ambiente proposto, pois a taxa de reconhecimento facial foi relativamente baixa, se comparada a outros trabalhos do estado da arte, trabalhos estes que usam de ambientes de testes mais amigáveis, mas ao mesmo tempo menos comumente encontrados em nosso dia-a-dia. Acredita-se que foi possível alcançar os objetivos propostos pelo trabalho e que o mesmo possa contribuir para o estado da arte atual na área de visão computacional, mais precisamente no âmbito do reconhecimento facial. Ao final são sugeridos alguns trabalhos futuros que podem ser utilizados como ponto de partida para a continuação desta pesquisa ou até mesmo de novas pesquisas relacionadas a este tema.

Palavras-chaves: Reconhecimento Facial. Detecção Facial. Baixa Resolução. Variação de Iluminação.

Abstract

Prado, Kelvin Salton do. **Detecting presence through face recognition under low resolution and low luminosity conditions**. 2018. 171 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2018. Corrected Version.

Face recognition is a task that human beings perform naturally in their everyday lives, usually with no effort at all. To machines, however, this process is not so simple. With the increasing computational power of current machines, a great interest was created in the field of digital videos and images processing, with applications in most diverse areas of knowledge. This work aims to compare face recognition techniques already know in the literature, in order to identify which technique has the best performance in a real and semicontrolled environment. As a secondary objective, we evaluate the possibility of using one or more face recognition techniques to automatically identify the presence of students in a martial arts classroom using images from the surveillance cameras installed in the room, taking into account important aspects such as images with low sharpness, illumination variation, constant movement of students and the fact that the cameras are at a fixed angle. This work is related to the Image Processing and Pattern Recognition areas, and integrates the research line “Presence Monitoring” of the project entitled “Education and Monitoring of Physical Activities using Artificial Intelligence Techniques.” (Process 2014.1.923.86.4, published in DOE 125 (45) on 03/10/2015), developed as a partnership between the University of São Paulo, Campo Limpo Paulista Faculty, and Kungfu-Wushu Central Academy. With the experiments performed and presented in this work it was possible to conclude that, amongst all face recognition methods that were tested, Local Binary Patterns had the best performance in the proposed environment. On the other hand, Eigenfaces had the worse performance according to the experiments. Moreover, it was also possible to conclude that it is not feasible to perform the automatic presence detection reliably in the proposed environment, since the face recognition rate was relatively low, compared to the state of the art which uses, in general, more friendly test environments but at the same time less likely found in our daily lives. We believe that it was possible to achieve the objectives proposed by this work and that can contribute to the current state of the art in the computer vision field and, more precisely, in the face recognition area. Finally, some future work is suggested that can be used as a starting point for the continuation of this work or even for new researches related to this topic.

Keywords: Face Recognition. Face Detection. Low Resolution. Illumination Variation.

Lista de figuras

Figura 1 – Exemplo de imagens de mesma dimensão (50×50 <i>pixels</i>)	22
Figura 2 – Imagens da academia Central Kungfu-Wushu	26
Figura 3 – Retângulos de características. Dois retângulos A e B, três retângulos C e quatro retângulos D.	38
Figura 4 – Exemplo do cálculo da imagem integral	39
Figura 5 – Esquema do classificador em cascata	40
Figura 6 – Face média calculada a partir da base de treinamento.	43
Figura 7 – Exemplo de operação básica do LBP.	46
Figura 8 – Exemplo de imagem resultante da operação básica do LBP.	46
Figura 9 – Extração do histograma da face.	47
Figura 10 – Exemplo de LBP circular.	48
Figura 11 – Procedimento de obtenção das Diferenças de Gaussianas (DoG) para diversas oitavas.	54
Figura 12 – Detecção de extremos no espaço-escala.	55
Figura 13 – Descrição dos pontos chave.	57
Figura 14 – Soma de intensidade dentro de uma região retangular em uma imagem integral.	60
Figura 15 – Gaussianas discretizadas e recortadas ($L_{yy}(a, \sigma)$ e $L_{xy}(a, \sigma)$) e as aproximações (D_{yy} e D_{xy}).	61
Figura 16 – Esquema de funcionamento do comitê de máquinas.	64
Figura 17 – Decomposição Wavelet 2D.	70
Figura 18 – Resultado da Decomposição Wavelet 2D.	71
Figura 19 – Possíveis resultados utilizando o limiar.	73
Figura 20 – Esquema da matriz de confusão utilizada pela curva ROC.	74
Figura 21 – Exemplo de interpolação utilizando o método do vizinho mais próximo.	77
Figura 22 – Exemplo de interpolação utilizando o método bilinear.	78
Figura 23 – Exemplo de interpolação utilizando o método bicúbico.	78
Figura 24 – Exemplo de imagem 1×10 <i>pixels</i>	78
Figura 25 – Exemplo dos principais métodos de interpolação.	79
Figura 26 – Exemplo dos principais métodos de interpolação.	80

Figura 27 – Esquema do processo de reconhecimento facial.	81
Figura 28 – Modelo de câmera utilizada nos experimentos.	82
Figura 29 – Imagens originais da ficha de cadastro da academia.	84
Figura 30 – Imagens faciais extraídas da ficha de cadastro da academia.	85
Figura 31 – Exemplo de imagens faciais utilizadas na sexta base de treinamento. . .	87
Figura 32 – Amostra das imagens faciais obtidas do próprio vídeo.	88
Figura 33 – Imagens faciais detectadas pelo algoritmo.	96
Figura 34 – Imagens não faciais detectadas pelo algoritmo.	96
Figura 35 – Exemplo das 3 salas com os respectivos números.	99
Figura 36 – Resultado de cada método de interpolação.	102
Figura 37 – Precisão média de cada algoritmo utilizando a primeira base de treina- mento.	110
Figura 38 – Precisão média de cada algoritmo utilizando a segunda base de treina- mento.	111
Figura 39 – Precisão média de cada algoritmo utilizando a terceira base de treinamento.	112
Figura 40 – Precisão média de cada algoritmo utilizando a quarta base de treinamento.	113
Figura 41 – Precisão média de cada algoritmo utilizando a quinta base de treinamento.	114
Figura 42 – Precisão média de cada algoritmo utilizando a sexta base de treinamento.	115
Figura 43 – Precisão média de cada algoritmo utilizando cada base de treinamento.	116
Figura 44 – Variação do limiar utilizando o método Eigenfaces.	120
Figura 45 – Variação do limiar utilizando o método LBPH.	121
Figura 46 – Variação do limiar utilizando o método Fisherfaces.	122
Figura 47 – Variação do limiar utilizando o método SIFT.	123
Figura 48 – Variação do limiar utilizando o método SURF.	124
Figura 49 – Curva ROC da comparação utilizando diferentes bases de treinamento.	128

Lista de quadros

Quadro 1 – Comparação das variáveis consideradas e das técnicas utilizadas . . .	36
Quadro 2 – Pontos positivos e negativos de cada método	63
Quadro 3 – Exemplo de comitê de máquinas utilizando voto majoritário	65
Quadro 4 – Exemplo de comitê de máquinas utilizando voto ponderado	67
Quadro 5 – Resumo das bases de treinamento	87
Quadro 6 – Resumo dos dados coletados da academia Central Kungfu-Wushu . . .	89
Quadro 7 – Parâmetros do algoritmo de detecção facial	92
Quadro 8 – Parâmetros do algoritmo de detecção facial	92
Quadro 9 – Parâmetros do algoritmo de detecção facial	93
Quadro 10 – Parâmetros do algoritmo de detecção facial	93

Lista de tabelas

Tabela 1 – Resumo dos testes realizados para a seleção dos parâmetros da etapa de detecção facial	95
Tabela 2 – Resumo dos resultados da operação de detecção facial em cada vídeo .	97
Tabela 3 – Resumo da quantidade de participantes detectados em cada vídeo. . .	98
Tabela 4 – Precisão dos algoritmos utilizando a primeira base de treinamento . . .	109
Tabela 5 – Precisão dos algoritmos utilizando a segunda base de treinamento . . .	110
Tabela 6 – Precisão dos algoritmos utilizando a terceira base de treinamento . . .	111
Tabela 7 – Precisão dos algoritmos utilizando a quarta base de treinamento . . .	112
Tabela 8 – Precisão dos algoritmos utilizando a quinta base de treinamento . . .	113
Tabela 9 – Precisão dos algoritmos utilizando a sexta base de treinamento . . .	114
Tabela 10 – Resumo da precisão e revocação dos experimentos envolvendo a base de treinamento 5 e o vídeo 5	126
Tabela 11 – Resumo da precisão e revocação dos experimentos envolvendo a base de treinamento 6 e o vídeo 11	127
Tabela 12 – Resultados do comitê de máquinas utilizando 3 componentes e voto majoritário	130
Tabela 13 – Resultados do comitê de máquinas utilizando 5 componentes e voto ponderado	131

Lista de abreviaturas e siglas

ACM	Association for Computing Machinery
API	Application Programming Interface
DoG	Difference of Gaussian
DWT	Discrete Wavelet Transform
FN	Falso Negativo
FP	Falso Positivo
FPR	False Positive Rate
HCI	Human Computer Interaction
IEEE	Institute of Electrical and Electronics Engineers
K-NN	K-Nearest Neighbors
LBP	Local Binary Patterns
LBPH	Local Binary Patterns Histogram
LDA	Linear Discriminant Analysis
LPP	Locality Preserving Projections
MBGC	Multiple Biometric Grand Challenge
MDS	Multidimensional Scaling
OpenCV	Open Source Computer Vision Library
PCA	Principal Component Analysis
RAISR	Rapid and Accurate Image Super Resolution
ROC	Receiver Operating Characteristic
ScFace	Surveillance Cameras Face Database
SIFT	Scale Invariant Feature Transform

SRC	Sparse Representation-based Classification
SURF	Speeded Up Robust Features
TPR	True Positive Rate
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
XML	Extensible Markup Language

Lista de símbolos

μ	Mi: representa a média das imagens faciais de treinamento.
Γ	Gama: vetor que representa cada imagem facial.
Φ	Fi: novo vetor que representa cada imagem facial, calculado pelo vetor médio (μ) subtraído do vetor linha da imagem (Γ).
λ	Lambda: representa os autovalores.
ω	Ômega: pesos calculados para as <i>eigenfaces</i> .
ϵ	Épsilon: distância euclidiana calculada no <i>eigenfaces</i> .
Ω	Ômega: vetor de pesos utilizado no <i>eigenfaces</i> .
σ	Sigma: representa a escala para os métodos SIFT e SURF.
θ	Teta: representa as orientações do gradiente no SIFT.

Sumário

1	Introdução	19
1.1	<i>Objetivos e Motivação do Trabalho</i>	20
1.1.1	Objetivo específico	22
1.2	<i>Visão geral do problema</i>	24
1.3	<i>Abordagem do problema</i>	27
1.4	<i>Organização do trabalho</i>	28
2	Trabalhos relacionados	29
2.1	<i>Revisão sistemática</i>	29
2.2	<i>Trabalhos selecionados</i>	30
2.3	<i>Considerações</i>	35
3	Fundamentação Teórica	37
3.1	<i>Detecção facial</i>	37
3.1.1	Algoritmo Viola-Jones	38
3.2	<i>Reconhecimento facial</i>	41
3.2.1	Eigenfaces	41
3.2.2	Local Binary Patterns Histograms (LBPH)	45
3.2.3	Fisherfaces	49
3.2.4	Scale Invariant Feature Transform (SIFT)	52
3.2.5	Speeded Up Robust Features (SURF)	58
3.2.6	Brute-Force Matcher	62
3.2.7	Comparação dos algoritmos	63
3.3	<i>Comitê de máquinas</i>	63
3.3.1	Voto majoritário	65
3.3.2	Voto ponderado	66
3.4	<i>OpenCV</i>	68
3.4.1	OpenCV contrib	68
3.5	<i>Transformada Wavelet</i>	69
3.6	<i>Curva ROC</i>	71
3.7	<i>Precisão e revocação</i>	75

3.8	<i>Interpolação Bicúbica</i>	77
4	Experimentos	81
4.1	<i>Informações técnicas</i>	82
4.2	<i>Dados coletados</i>	83
4.2.1	Academia Central Kungfu-Wushu	83
4.2.2	Participantes da pesquisa	83
4.2.3	Bases de treinamento	84
4.2.4	Bases de teste	89
4.2.5	Questões éticas	90
4.3	<i>Deteção facial</i>	90
4.3.1	Descrição dos parâmetros	90
4.3.2	Seleção dos parâmetros	91
4.3.3	Resultados da deteção facial	97
4.4	<i>Pré-processamento dos dados</i>	100
4.4.1	Redimensionamento das Imagens	101
4.5	<i>Reconhecimento facial</i>	103
4.5.1	Descrição dos parâmetros	103
4.5.2	Resultados principais	107
4.5.3	Resultados variando o limiar	117
4.5.4	Resultados do comitê de máquinas	128
5	Considerações finais	133
6	Conclusões	137
6.1	<i>Trabalhos futuros</i>	138
	Referências¹	140
	Apêndice A – Protocolo da Revisão Sistemática	146
	Apêndice B – Condução da Revisão na Base da ACM	150
	Apêndice C – Condução da Revisão na Base da IEEE	161

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

Apêndice D – Termo de Consentimento Livre e Esclarecido .	166
Apêndice E – Termo de Assentimento	167
Apêndice F – Resultados dos Testes com os Diversos Métodos de Interpolação	169

1 Introdução

O reconhecimento facial é uma tarefa que as pessoas realizam automaticamente todos os dias em suas vidas, e praticamente sem esforço nenhum. Entretanto, para uma máquina o processo de reconhecimento facial não é uma tarefa tão simples, uma vez que há uma série de variáveis envolvidas (JAIN; LI, 2005).

Até meados dos anos 90 as técnicas de reconhecimento facial eram relativamente simples, sendo a maior parte dos trabalhos nesta área focados em realizar o reconhecimento de apenas uma face com um fundo simples ou complexo (ZHAO et al., 2003). Porém, com o aumento do poder computacional das máquinas atuais, a área de reconhecimento facial cresceu muito, desenvolvendo-se novas técnicas e métodos, os quais tornaram possível um reconhecimento mais preciso mesmo em ambientes arbitrários. Tendo em vista estas perspectivas, criou-se um enorme interesse no processamento automático de imagens e vídeos digitais, com aplicações nas mais diversas áreas de conhecimento.

De acordo com Jain e Li (2005), em um sistema de reconhecimento facial espera-se reconhecer faces presentes em imagens ou vídeos de forma automática, podendo o mesmo operar de dois modos:

1. **Verificação ou autenticação da face:** consiste na comparação de 1 para 1, ou seja, comparar uma face que está sendo consultada com outra face para a qual está sendo requerida a autenticação.
2. **Identificação ou reconhecimento de face:** consiste na comparação de 1 para N, ou seja, compara uma imagem facial consultada em relação a todas as imagens faciais da base de dados, para determinar a identidade da face consultada.

No desenvolvimento deste trabalho é utilizada a identificação ou reconhecimento de face, sendo que para cada face localizada no vídeo a mesma é comparada com todas as faces da base de dados para determinar a identidade do sujeito.

Este trabalho faz parte da linha de pesquisa de “Monitoramento de Presença” do projeto “Ensino e Monitoramento de Atividades Físicas via Técnicas de Inteligência Artificial” (Processo 2014.1.923.86.4, publicado no DOE 125(45), em 10/03/2015), realizado conjuntamente pela Universidade de São Paulo, Faculdade Campo Limpo Paulista e Academia Central Kungfu-Wushu, tratando-se então de um trabalho teórico e prático que será implementado no ambiente real da academia de artes marciais.

Dentro do projeto “Ensino e Monitoramento de Atividades Físicas via Técnicas de Inteligência Artificial”, o presente trabalho tem como objetivo específico analisar a possibilidade de realizar a detecção de presença de forma automática por meio do uso de técnicas de reconhecimento facial.

Sendo assim, ao final do trabalho pretende-se identificar, dentre os métodos de reconhecimento facial estudados, qual método tem o melhor desempenho no ambiente real ao qual ele será aplicado, bem como concluir sobre a possibilidade de utilizar o reconhecimento facial para realizar a detecção de presença de pessoas sob as mesmas condições que as encontradas em nosso ambiente de teste. Os objetivos gerais e específicos serão mais bem detalhados nas seções seguintes.

1.1 Objetivos e Motivação do Trabalho

O objetivo geral deste trabalho encontra-se na comparação de técnicas de reconhecimento facial, já conhecidas na literatura, em um ambiente real, semicontrolado e não colaborativo, visando a identificar a técnica que apresenta melhores resultados nesse ambiente, levando em consideração as diversas variáveis envolvidas, tais como a variação de iluminação e a baixa nitidez das imagens.

A principal motivação do trabalho está relacionada ao ambiente arbitrário no qual pretende-se realizar o reconhecimento facial. A maior parte dos trabalhos na área de reconhecimento facial são realizados em ambientes controlados (e.g. (BISWAS et al., 2013; DINIZ et al., 2013; MAO; LI; YIN, 2014; XU; LIU; LI, 2014)) e de forma colaborativa, onde os participantes são instruídos a registrarem suas imagens faciais em posições pré-definidas, geralmente com uma boa resolução e iluminação ou até mesmo utilizando bases de imagens já criadas nestas circunstâncias.

A abordagem citada faz sentido quando aplicada à área de biometria, pois neste caso o usuário está colaborando com o sistema e pode ser instruído a registrar a imagem facial de uma forma específica, auxiliando de certa forma o sistema a realizar a autenticação facial de forma mais precisa e confiável.

Todavia, quando o objetivo da aplicação é realizar o reconhecimento facial em um ambiente real, semicontrolado e não colaborativo, começam a surgir diversas variáveis que podem prejudicar a precisão dos métodos de reconhecimento facial, tais como baixa

resolução, variação de iluminação, oclusão, dentre outras. O presente trabalho aborda exatamente este ponto, comparando os métodos de reconhecimento facial, já existentes na literatura, em um ambiente real com o objetivo de avaliar a possibilidade de utilização de tais métodos em cenários reais nos quais não há, geralmente, uma boa base de treinamento¹ e a qualidade das imagens é relativamente baixa.

Cabe esclarecer que o termo “semicontrolado” é utilizado neste trabalho pelo fato de não se ter um controle total em relação ao ambiente de estudos, ou seja, a sala de aula da academia. Em geral, este fato ocorre principalmente por causa dos problemas relacionados ao ambiente (e.g. variação de iluminação, oclusão) e pelo fato de que os alunos e instrutores se movimentam livremente pelo recinto, sem ter o intento de ter a face capturada pelas câmeras, o que torna este ambiente não colaborativo também.

Para o cumprimento dos objetivos do trabalho, bem como uma melhor compreensão destes, convém esclarecer que, devido à variação da resolução das imagens², usaremos neste trabalho o termo “nitidez” para representar qualidade das imagens faciais obtidas dos vídeos. Ainda, será utilizado o termo “dimensão” para representar o tamanho das imagens em *pixels*. Desta forma, as imagens faciais extraídas dos vídeos possuem uma dimensão média de 42×42 *pixels* e baixa nitidez, se comparadas às imagens faciais obtidas da ficha de cadastro dos alunos. Ainda, as imagens faciais são originalmente extraídas nesta dimensão e qualidade, diferenciando-se de outros trabalhos nos quais as imagens faciais são obtidas de câmeras de alta resolução, sendo então aplicados métodos de reescalonamento para diminuir a qualidade das imagens. Neste último caso, apesar de a dimensão da imagem ser considerada baixa, a qualidade da imagem é superior, se comparada a outras imagens de baixa qualidade obtidas, por exemplo, de câmeras de vigilância. Como exemplo, a Figura 1 representa duas imagens faciais de dimensão 50×50 *pixels* da mesma pessoa. A imagem facial da esquerda foi obtida diretamente de um vídeo da academia e a imagem facial da direita foi obtida da ficha de cadastro do aluno e reescalada para ficar com a mesma dimensão.

¹ Considera-se que uma boa base de treinamento deve ser composta por diversas imagens de um mesmo sujeito em diferentes condições.

² A variação de resolução está relacionada principalmente à distância entre os sujeitos e as câmeras, sendo que quanto mais próximo da câmera o sujeito estiver melhor é a resolução, pois existem mais *pixels* por polegadas, mesmo após o redimensionamento da imagem.

Figura 1 – Exemplo de imagens de mesma dimensão (50×50 pixels)

Fonte: Kelvin Salton do Prado, 2017

Como é possível visualizar, apesar de as duas imagens faciais possuírem as mesmas dimensões, a nitidez delas é bem diferente, o que acaba dificultando um processo de reconhecimento facial mais preciso.

Apesar das dificuldades encontradas em ambientes reais, acredita-se que os estudos e as aplicações nestes ambientes são de suma importância, pois possibilitam realizar desde uma detecção de presença de alunos em sala de aula até a identificação de um suspeito em uma multidão.

Tendo em vista o exposto acima, o presente trabalho tem como motivação realizar a comparação de diversas técnicas de reconhecimento facial em um ambiente real, semi-controlado e não colaborativo, com o objetivo primário de identificar qual das técnicas estudadas tem melhor desempenho no ambiente proposto, bem como o objetivo secundário de analisar a possibilidade de utilização de métodos de reconhecimento facial para a detecção automática de pessoas nesse tipo de ambiente.

1.1.1 Objetivo específico

Com o propósito de facilitar a compreensão em relação ao objetivo específico deste trabalho, cabe esclarecer que a identificação dos alunos e instrutores da academia Central Kungfu-Wushu é atualmente realizada por meio de fichas anotadas e contabilizadas de forma manual. Esse tipo de controle, contudo, apresenta algumas dificuldades específicas, como o fato dos alunos que estão na academia não estarem necessariamente em aula (especialmente em vista de que alguns vão para treinos extraclasse) e a falta de um controle mais rigoroso da presença dos instrutores durante as aulas.

Com isto, propõe-se como objetivo específico o estudo da viabilidade do uso de técnicas de reconhecimento facial já conhecidas na literatura, aplicadas ao ambiente

semicontrolado e não colaborativo da academia, para identificação automática da presença de alunos e instrutores durante as aulas.

Neste sentido, surgem algumas dificuldades relacionadas ao ambiente semicontrolado da sala de aula de artes marciais, tais como a câmera estar em uma posição fixa em um canto da sala, o problema da oclusão, a baixa nitidez das imagens, a luminosidade não ideal e a variação de pose (esses problemas serão mais bem discutidos na Seção 1.2). Nestes ambientes, o reconhecimento facial ainda encontra diversos desafios que podem prejudicar significativamente seu desempenho.

Além das dificuldades citadas, há também o fato de que, durante uma aula, os alunos podem ser separados em sub-grupos menores, que por sua vez podem ser separados em salas distintas. Nesse caso, embora participando da mesma aula e mesma classe, cada sub-grupo é capturado por uma câmera diferente, dificultando ainda mais a identificação do aluno em relação a sua turma. Como agravante, a nova sala que recebe o sub-grupo pode já conter um grupo de alunos. Assim sendo, as seguintes configurações são possíveis:

- Alunos de classes distintas em salas distintas, sem compartilhamento do mesmo espaço físico.
- Alunos de classes distintas compartilhando uma mesma sala, caso em que a sala é fisicamente sub-dividida entre os grupos ou sub-grupos.
- Alunos de classes distintas, separados em sub-grupos, compartilhando uma ou mais salas, caso em que a sala também é fisicamente sub-dividida, mas as classes nelas contidas não estão completas.

A ausência dos alunos durante certo período de tempo é um problema significativo para o objetivo secundário deste trabalho, uma vez que os alunos não estão na sala que deveriam estar. Por outro lado, tal ausência não interfere no objetivo principal deste trabalho, pois a comparação dos algoritmos de reconhecimento facial será realizada utilizando apenas as imagens obtidas, não interferindo portanto no fator comparação. Tendo isto em vista, acredita-se que este problema está relacionado à regra de negócio e não deve ser considerado como uma variável do ambiente que possa ser mais bem estudada e tratada.

Sendo assim, acredita-se que os problemas mencionados nesta seção associados ao problema real encontrado na academia alvo do estudo, o qual envolve a falta de um controle mais rigoroso em relação à presença dos alunos em sala de aula, são grandes

desafios para o bom desempenho de um sistema de reconhecimento facial, motivando ainda mais o desenvolvimento deste trabalho.

Além disto, como já mencionado anteriormente, salienta-se que o projeto de pesquisa proposto pode ser utilizado futuramente como base para outros trabalhos relacionados à área de reconhecimento facial em ambiente reais, não se limitando apenas ao ambiente no qual o trabalho será aplicado.

Na seção seguinte será apresentada uma visão geral do problema e das variáveis envolvidas no ambiente real da academia.

1.2 Visão geral do problema

Cabe aqui esclarecer novamente e resumidamente o objetivo geral e específico deste trabalho:

- Objetivo geral: comparar diversas técnicas de reconhecimento facial em um ambiente real, semicontrolado e não colaborativo com o objetivo de identificar a técnica que melhor se adapte ao ambiente estudado.
- Objetivo específico: analisar a possibilidade de realizar uma detecção de presença automática, em um ambiente real, utilizando uma ou mais das técnicas de reconhecimento facial estudadas.

Para a execução do objetivo específico foi realizado um estudo em uma academia de artes marciais. Nesse caso, um dos principais problemas encontrados atualmente pela academia de artes marciais é a falta de um controle mais rigoroso na identificação dos alunos e instrutores durante as aulas. Observou-se que muitas vezes os alunos esquecem a carteira de identificação para acessar o ambiente da academia, sendo necessário um funcionário da academia liberar o acesso ao aluno, perdendo assim, de certa forma, o controle sobre os alunos que a estão frequentando. Além deste problema, notou-se que alguns alunos frequentam a academia para realizar treinos extraclasse, o que também prejudica o controle ideal.

Pelos motivos apresentados, propõe-se o estudo e implementação de uma ou mais técnicas de reconhecimento facial com o intuito de realizar a detecção de presença de alunos e instrutores na academia, utilizando para tal imagens obtidas das câmeras de segurança do local.

Por se tratar de um ambiente real, o reconhecimento facial pode ser afetado por diversos aspectos, os quais podem prejudicar significativamente o seu desempenho. Tais aspectos envolvem:

- **Câmera em ângulo fixo:** foram utilizadas imagens das câmeras de segurança do local, as quais já estavam instaladas no recinto e possuem um ângulo fixo. Cabe esclarecer que as câmeras se encontram em pontos estratégicos com o objetivo de capturar imagens amplas das salas. Sendo assim, não estão em um posicionamento ideal para o processo de reconhecimento facial. Neste caso, entende-se como posicionamento ideal um local onde as câmeras estariam na altura da face dos sujeitos e que pudessem capturar uma imagem frontal dela. Ainda, cabe salientar que alterações físicas no ambiente da academia, como por exemplo alterar a posição das câmeras, não fazem parte do escopo deste trabalho.
- **Baixa nitidez:** por se tratarem de câmeras de segurança comuns, a nitidez das imagens é relativamente baixa.
- **Variação de iluminação:** constatou-se que o ambiente da sala de aula analisada nesse trabalho sofre variações de iluminação, principalmente referentes a determinados horários do dia. De acordo com Adini, Moses e Ullman (1997), a variação de iluminação é um dos fatores críticos que afeta drasticamente a performance do reconhecimento facial.
- **Movimentação constante:** por se tratar de uma academia de artes marciais, os alunos geralmente estão em constante movimentação, e isto pode prejudicar a qualidade das imagens capturadas, além de gerar outros dois problemas: a variação de pose e a oclusão.
- **Variação de pose:** como os alunos estão em constante movimentação durante as aulas, ocorre de as imagens faciais capturadas terem variação de pose, o que não é ideal para o bom desempenho do reconhecimento facial. Cabe esclarecer que a expressão “variação de pose” refere-se às variações na posição e ângulo da face capturada nas imagens.
- **Oclusão:** o problema da oclusão pode ocorrer quando a face de um sujeito estiver sendo ocultada por outro sujeito ou objeto, ou quando o próprio sujeito estiver ocultando a sua face ou parte dela com outras partes do corpo (e.g. cabelo, mão) (BAK et al., 2012; WEI et al., 2014). Entende-se que o problema da oclusão pode ocorrer

devido à movimentação constante da pessoa, bem como, pelo fato da câmera estar em um ângulo fixo. Porém, acredita-se que esse problema pode ser mitigado, uma vez que com a movimentação constante a oclusão poderá ocorrer em alguns *frames*³ mas não em outros.

Cabe esclarecer que, apesar de compreender que os problemas citados anteriormente são grandes desafios para o bom desempenho da tarefa de reconhecimento facial, não se tem como objetivo deste trabalho tratar estes problemas diretamente. Apesar disso, serão utilizados métodos de reconhecimento facial que são considerados, pela literatura relacionada, eficientes para lidar com alguns destes problemas, tais como variação de iluminação e variação de pose, amenizando assim os efeitos causados pelos mesmos.

Para ter uma ideia geral sobre o ambiente no qual o trabalho se insere, a Figura 2 apresenta quatro imagens obtidas da academia Central Kungfu-Wushu, representando o ambiente real das salas de aula da academia. Como é possível notar nas imagens, em todas as salas a câmera se encontra em uma posição fixa e, de forma geral, desfavorável para o processo de reconhecimento facial, muitas vezes capturando imagens dos sujeitos de costas, o que exemplifica as dificuldades apresentadas anteriormente.

Figura 2 – Imagens da academia Central Kungfu-Wushu

Fonte: Kelvin Salton do Prado, 2016

³ A expressão *frames* origina-se da língua inglesa e pode ser traduzida para o português como *quadros*. Optou-se por manter a palavra na língua inglesa pois ela é comumente utilizada em aplicações e trabalhos que utilizam vídeos.

Embora diversos trabalhos a respeito de reconhecimento facial já tenham sido desenvolvidos, poucos tratam de ambientes não controlados (e.g. (BISWAS et al., 2013; MAO; LI; YIN, 2014)), onde as imagens possuem baixa qualidade, há variação de luminosidade ou as pessoas encontram-se em constante movimentação. Cabe aqui ressaltar que até o momento não foram encontrados outros trabalhos envolvendo todos os aspectos citados neste trabalho, nem mesmo em condições tão desfavoráveis para um reconhecimento facial preciso. Na próxima seção será apresentada uma abordagem do problema, buscando um melhor entendimento do trabalho como um todo.

1.3 Abordagem do problema

Buscando amenizar os problemas, semelhantes aos encontrados na academia alvo deste trabalho, propõe-se o estudo e a comparação de técnicas de reconhecimento facial, com o objetivo de identificar o método que apresentar o melhor resultado no ambiente estudado. Tendo em vista que pretende-se utilizar o reconhecimento facial para identificar pessoas em um ambiente real e semicontrolado, pode-se então subdividir este processo em duas etapas básicas:

1. **Detecção da face:** dada uma imagem qualquer (neste caso um *frame* do vídeo) o algoritmo detecta as faces na imagem, sendo possível extrair múltiplas faces de uma mesma imagem.
2. **Reconhecimento da face:** cada imagem facial encontrada é comparada com as demais faces da base de dados, sendo então apontado o sujeito da imagem, com um determinado grau de confiança.

Além disto, propõe-se uma abordagem de comitê de máquinas, descrita na Seção 3.3, com o objetivo de melhorar os resultados obtidos em relação às técnicas de reconhecimento facial independentes.

Cabe enfatizar que, para a realização da detecção facial, neste trabalho foi utilizada a técnica de Viola-Jones, descrita na Seção 3.1.1, por ser uma técnica bastante difundida na literatura (e.g. (DINIZ et al., 2013; ARAUJO, 2010)). Apesar de a detecção facial ser uma etapa importante no processo de detecção de presença, não a temos como foco principal deste trabalho.

No processo de reconhecimento facial foram utilizadas as seguintes técnicas: *Eigenfaces*, *Fisherfaces*, *Local Binary Patterns Histogram (LBPH)*, *Scale-Invariant Feature Transform (SIFT)* e *Speeded Up Robust Features (SURF)*. Estas técnicas foram selecionadas por serem bem difundidas na área de processamento de imagens e reconhecimento facial (e.g. (AHONEN; HADID; PIETIKÄINEN, 2004; AMARAL; GIRALDI; THOMAZ, 2013; MACHADO et al., 2009; DINIZ et al., 2013; SOUZA, 2014)), por serem disponibilizadas pela biblioteca OpenCV⁴ e serem, de certa forma, robustas para lidar com alguns dos problemas citados, como por exemplo a variação de iluminação e variação de pose (LOWE, 1999; BAY; TUYTELAARS; GOOL, 2006; MACHADO et al., 2008; OPENCV, 2000). As técnicas e abordagens utilizadas serão conceituadas em detalhes nos capítulos seguintes.

1.4 Organização do trabalho

O restante desse trabalho está organizado como segue. No Capítulo 2 serão apresentados alguns trabalhos relacionados que envolvem o processo de reconhecimento facial e que abordam alguns dos problemas citados nas seções anteriores. O objetivo deste capítulo é de familiarizar o leitor com algumas obras da literatura e fazer uma breve comparação com o trabalho proposto, bem como com as técnicas e variáveis consideradas. Em seguida, o Capítulo 3 apresenta uma fundamentação teórica de todas as técnicas e abordagens utilizadas, buscando facilitar o entendimento do trabalho como um todo.

Por sua vez, o Capítulo 4 apresenta todos os experimentos realizados e os resultados obtidos. Neste capítulo são descritos os passos realizados e os parâmetros utilizados nos experimentos, buscando detalhá-los para que possam futuramente ser reproduzidos por outros pesquisadores. Além disso, esse capítulo apresenta os resultados obtidos com cada experimento, com a intenção de comprovar qual das técnicas estudadas teve o melhor desempenho no ambiente proposto, bem como comprovar a possibilidade ou não de se utilizar as técnicas de reconhecimento facial com o objetivo de realizar a identificação de alunos e instrutores da academia.

Por fim, os Capítulos 5 e 6 apresentam algumas considerações e conclusões obtidas, bem como alguns trabalhos futuros que poderão ser realizados.

⁴ A biblioteca OpenCV (Open Source Computer Vision Library) é descrita com mais detalhes na Seção 3.4.

2 Trabalhos relacionados

Para melhor compreender o propósito geral deste trabalho, e ainda, o âmbito no qual ele está inserido, foram selecionados alguns trabalhos relacionados, apresentados nas seções seguintes. Esses trabalhos foram escolhidos por tratarem de situações semelhantes às propostas por esta pesquisa, pois acredita-se que as funções e técnicas utilizadas são relevantes para a finalidade de comparação de condições de contorno, objetivando ainda apresentar as principais semelhanças e diferenças entre o trabalho desenvolvido e os trabalhos relacionados.

2.1 Revisão sistemática

Visando a realizar um levantamento do estado da arte atual, bem como localizar as principais técnicas utilizadas, foi realizada uma revisão sistemática, englobando principalmente trabalhos relacionados ao reconhecimento facial levando em consideração o problema da variação de iluminação. A revisão sistemática consiste em uma metodologia científica que vai além de uma simples visão geral do estado da arte. Sendo assim, trata-se de uma pesquisa capaz de identificar, selecionar e produzir provas em relação a um determinado assunto (MIAN et al., 2005).

Inicialmente foi realizado um estudo exploratório nas bases de dados da IEEE Xplore¹ e do Portal ACM², com o objetivo de levantar os dados necessários para a geração do protocolo de revisão, o qual pode ser encontrado em sua totalidade no Apêndice A. O protocolo de revisão tem o objetivo de detalhar como a revisão sistemática será conduzida, apresentando, por exemplo, o objetivo da pesquisa, as palavras-chave e os critérios de inclusão e exclusão dos artigos.

Durante a condução da revisão foram realizadas buscas nessas bases de dados utilizando *strings* de buscas criadas a partir das palavras-chaves definidas no protocolo de revisão. As buscas tinham como objetivo principal localizar trabalhos de língua inglesa e, por este motivo, foram utilizadas as seguintes palavras-chave: *face recognition*, *face detection*, *low resolution*, *face tracking*, *surveillance camera*, *illumination invariance*, *pose invariant* e *person recognition*.

¹ Site da IEEE Xplore: <http://ieeexplore.ieee.org>

² Site do Portal ACM: <http://dl.acm.org/>

A primeira busca, realizada na base de dados da ACM, retornou 77 artigos, dos quais apenas 10 foram definidos como satisfatórios de acordo com os critérios de inclusão. Enquanto na segunda busca, realizada na base de dados da IEEE, foi possível localizar 27 artigos, dos quais apenas 6 foram incluídos de acordo com os critérios estabelecidos. Mediante a leitura dos artigos incluídos foi possível realizar uma nova filtragem e excluir mais um artigo, restando, portanto, ao final, 15 artigos incluídos. Os protocolos de condução da revisão da base de dados da ACM e da IEEE podem ser visualizados na íntegra nos Apêndices B e C respectivamente.

Por meio da leitura dos artigos foi possível identificar que os problemas de variação de iluminação, variação de pose e baixa resolução são reais e afetam significativamente o desempenho do reconhecimento facial. Com isto, foi possível observar também algumas estratégias adotadas para minimizar o problema da variação de iluminação como, por exemplo, a utilização de algoritmos de “re-iluminação” (e.g. (ARANDJELOVIĆ; CIPOLLA, 2006; ARANDJELOVIĆ; CIPOLLA, 2013)) e o uso de imagens termais (e.g. (ARANDJELOVIĆ; HAMMOUD; CIPOLLA, 2010; HEO; SAVVIDES; VIJAYAKUMAR, 2005)). Além disto, os autores utilizam diversas técnicas conhecidas na literatura, tais como a Análise de Componentes Principais (PCA) e o algoritmo de Viola-Jones (e.g. (ARANDJELOVIC; CIPOLLA, 2008; ARANDJELOVIĆ; CIPOLLA, 2013)). Ao final, foi desenvolvido e publicado o artigo denominado “*Automatic Facial Recognition: A Systematic Review on the Problem of Light Variation*” (PRADO et al., 2016), apresentado na conferência internacional em interação humano-computador (HCI International) no ano de 2016.

Cabe esclarecer que, como a revisão sistemática tinha o objetivo de localizar trabalhos que envolviam o reconhecimento facial levando em consideração principalmente o problema da variação de iluminação, e que a mesma foi conduzida apenas em bases de dados de língua estrangeira (IEEE e ACM), tal revisão foi utilizada principalmente como base para a seleção dos métodos de reconhecimento facial utilizados neste trabalho, e os trabalhos relacionados, apresentados nas seções seguintes, não são em sua totalidade derivados dela.

2.2 Trabalhos selecionados

O primeiro trabalho selecionado, proposto por Mao, Li e Yin (2014), tem como objetivo principal detectar, rastrear e reconhecer alunos em uma sala de aula real, com o

propósito de identificar a presença dos mesmos, utilizando para isto métodos de reconhecimento facial. Os autores citam alguns problemas relacionados à variação da expressão facial, variação de pose, condições de iluminação, oclusão, entre outros, que prejudicam a performance do reconhecimento facial. A câmera utilizada no trabalho está alocada juntamente com o instrutor, o que possibilita sua movimentação e, neste caso, apesar de ocorrer o problema da oclusão, a mesma é minimizada pela possibilidade de movimentação da câmera.

Utilizando um detector facial em cascata, baseado em características *Haar*, os autores primeiramente detectam as faces no vídeo para posteriormente realizar o reconhecimento facial utilizando o método *eigenfaces*, considerando uma base de treinamento com imagens de cada aluno em diversas posturas, exigindo assim uma aquisição prévia das imagens para treinamento.

Com o objetivo de melhorar a acurácia, os autores utilizam uma abordagem de rastreamento, criando grupos de imagens do mesmo sujeito que posteriormente são utilizados para a realização do reconhecimento facial, o que provou melhorar os resultados se comparado ao reconhecimento facial em cada imagem independente. Os testes foram realizados utilizando a base de dados Honda/UCSD³ e também o cenário real da sala de aula.

Os autores concluem que, com os testes utilizando a base de dados Honda/UCSD, o processo de reconhecimento é mais preciso quando se utiliza a abordagem de criação de grupos de imagens do mesmo sujeito, onde foi possível obter acurácias de até 100%, se comparado ao reconhecimento facial em uma única imagem, onde foi possível obter acurácia máxima de 93% e média de 75%. Concluem ainda que, devido ao alto desempenho do reconhecimento obtido utilizando o cenário real da sala de aula, a verificação de frequência dos alunos baseada em vídeo se mostra promissora e viável.

O trabalho relacionado assemelha-se a esta pesquisa, pois também trata de um ambiente real, envolvendo problemas semelhantes, tais como a oclusão, a variação de pose e as condições de iluminação não ideais.

As principais diferenças entre o trabalho relacionado e esta pesquisa são:

³ Honda/UCSD: vision.ucsd.edu/~leekc/HondaUCSDVideoDatabase/HondaUCSD.html

- Nesta pesquisa, como trata-se de uma sala de aula de artes marciais, os alunos estão em constante movimentação, fator que difere do trabalho relacionado, e que dificulta ainda mais o reconhecimento facial preciso;
- Ainda, as câmeras utilizadas para obtenção das imagens estão em uma posição fixa e em um ângulo não favorável para a realização do reconhecimento facial, enquanto no trabalho relacionado a câmera é móvel e está localizada com o instrutor.

O segundo trabalho relacionado, desenvolvido por Biswas et al. (2013), propõe uma nova abordagem para combinar imagens faciais com qualidade de câmeras de vigilância em relação a imagens de alta resolução de pose frontal. Os autores esclarecem que as câmeras de segurança geralmente possuem baixa resolução⁴, e com a adição de variação de pose e de iluminação isto representa um problema, afetando o desempenho dos algoritmos de reconhecimento facial. Assim, os algoritmos de reconhecimento facial têm melhor desempenho quando as imagens capturadas e as imagens do banco de dados possuem a mesma qualidade e são capturadas em condições similares ou controladas.

Com o objetivo de amenizar este problema, os autores propõem o desenvolvimento de uma nova abordagem, chamada *Multidimensional Scaling (MDS)*, a fim de combinar imagens faciais de baixa resolução, obtidas das câmeras de segurança, com imagens de alta resolução de pose frontal, obtidas da base de dados.

Os autores utilizam técnicas como SIFT, PCA e *Tensor Analysis*, entre outras, a fim de encontrar os atributos da face na imagem e assim normalizá-la em relação à postura, para então realizar a comparação com a base de imagens.

Os autores utilizaram as bases Multi-PIE database⁵ e Surveillance Cameras Face Database (ScFace)⁶, bem como os dados obtidos do Multiple Biometric Grand Challenge (MBGC)⁷ nos experimentos. Ao final, concluem que, a abordagem proposta supera o estado da arte de técnicas de super-resolução baseadas em classificadores. Além disto, uma nova abordagem baseada em análise de tensores⁸ é proposta para realizar a estimativa automática de pose.

A principal diferença entre o trabalho proposto pelos autores e esta pesquisa é que enquanto o primeiro tem o foco principal na resolução das imagens faciais, buscando

⁴ O termo “baixa resolução” (do inglês *low-resolution*) é empregado pelos autores para representar imagens de baixa dimensão e, consequentemente, com menor nitidez.

⁵ Multi-PIE: <http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Home.html>

⁶ ScFace: <http://www.scface.org/>

⁷ MBGC: <https://www.nist.gov/programs-projects/multiple-biometric-grand-challenge-mbgc>

⁸ Do inglês *tensor analysis*.

principalmente aproximar imagens de baixa resolução a imagens de alta resolução, o segundo tem como objetivo a comparação de métodos de reconhecimento facial em um ambiente real, levando em consideração alguns aspectos, tais como a variação de iluminação e a oclusão, os quais podem prejudicar significativamente o desempenho do reconhecimento facial. Ainda assim, acredita-se que este trabalho relacionado apresenta informações relevantes que podem agregar valor à pesquisa desenvolvida, principalmente em relação à qualidade das imagens de treinamento em comparação às imagens de teste.

O terceiro trabalho relacionado, proposto por Diniz et al. (2013), tem como objetivo principal aplicar o reconhecimento facial na identificação de alunos que acessam um ambiente virtual de aprendizagem. Por conseguinte, espera-se que o aluno esteja na maior parte do tempo de frente para a câmera, pois trata-se de uma *webcam* instalada em um computador.

A partir disto, os autores realizaram vários testes com diversas técnicas e algoritmos para identificar qual se mostrava melhor para o ambiente proposto pelo trabalho. Dentre as técnicas utilizadas, pode-se citar como exemplo o método Eigenfaces, para a extração de características da face, e os classificadores K-Nearest Neighbors (K-NN), Random Forest e K-Star no processo de reconhecimento facial.

Os testes e validações dos algoritmos foram realizados em uma base de dados própria, contendo 1280 imagens de 64 classes distintas, sendo que cada indivíduo integrante desta base é representado por 20 imagens em variadas posições. Os autores concluem que todos os três classificadores testados tiveram um bom desempenho, chegando a atingir acurácias acima de 90% com determinados parâmetros, sendo que o melhor resultado foi obtido utilizando o classificador *K-NN*, com o qual foi possível obter uma taxa de reconhecimento de aproximadamente 92%.

Ainda que esta pesquisa se diferencie em vários aspectos do trabalho proposto pelo autor, pois seu objetivo principal é a aplicação do reconhecimento facial em um ambiente virtual de aprendizagem, não sendo necessária uma preocupação maior com fatores como variação de iluminação, variação de pose e oclusão, ele foi aqui apresentado por tratar da utilização de diferentes técnicas de reconhecimento facial, as quais podem ser utilizadas com a finalidade de comparação com o trabalho proposto.

O principal objetivo do quarto trabalho relacionado, proposto por Xu, Liu e Li (2014), consiste em melhorar o desempenho do reconhecimento facial em um ambiente

real, utilizando câmeras de segurança, levando em consideração três fatores importantes que influenciam este desempenho:

1. O tipo de câmera utilizada;
2. A distância entre a câmera e o sujeito; e
3. A resolução das imagens faciais capturadas.

De acordo com Xu, Liu e Li (2014), apenas uma pequena parte das pesquisas existentes nesta área são especificamente realizadas em ambientes reais, onde as faces capturadas são bem diferentes das faces obtidas em um ambiente controlado. Os autores ainda esclarecem que a maior parte dos trabalhos utilizando imagens de baixa resolução baseiam-se no que eles (Xu *et. al.*) definem como *down-sampled low-resolution*, representando imagens faciais obtidas originalmente de câmeras de alta resolução, porém submetidas a métodos de reescalonamento e/ou suavização. Além disso, os autores concluem que imagens de baixa resolução geradas “artificialmente” não são boas substitutas para imagens diretamente capturadas em baixa resolução. Cabe esclarecer ainda que os autores utilizaram técnicas para amenizar o problema da variação de iluminação com a intenção de melhorar o desempenho do reconhecimento facial.

Em seus experimentos, foram utilizadas 4 bases de imagens: FRGC⁹, AR¹⁰, ScFace¹¹ e Curtin Faces¹², sendo que a FRGC e AR compreendem imagens capturadas com câmeras de alta resolução, enquanto as bases ScFace e Curtin Faces possuem tanto imagens de alta resolução quanto imagens de câmeras de segurança. A partir dessas imagens, os autores utilizaram então os métodos Análise Discriminante Linear (LDA), *Locality Preserving Projections* (LPP), Análise de Componentes Principais (PCA) e *Sparse Representation-based Classification* (SRC) para a realização do reconhecimento facial.

De acordo com os experimentos, a resolução das imagens foi considerada o principal fator que afeta o desempenho do reconhecimento facial em ambientes que utilizam câmeras de vigilância, na medida em que quanto maior a resolução das imagens, melhor é o desempenho do reconhecimento facial. Os autores colocam ainda que a abordagem proposta supera as abordagens tradicionais de reconhecimento facial, mas esclarecem que, apesar dos testes terem um bom desempenho em um ambiente interno (*indoor*) e com imagens

⁹ FRGC: www.nist.gov/itl/iad/ig/frgc.cfm

¹⁰ AR: www2.ece.ohio-state.edu/~aleix/ARdatabase.html

¹¹ ScFace: www.scface.org

¹² Curtin Faces: não foi possível localizar o *link* para esta base.

de baixa resolução, ainda é preciso realizar mais testes em ambientes externos (*outdoor*) e com distâncias mais longas em relação à posição da câmeras e dos sujeitos.

Além disso, os autores esclarecem que a diferença entre câmeras e resoluções levam a significativas diferenças na taxa de reconhecimento, sendo que, com uma distância de 5 metros entre os sujeitos e a câmera, foi possível obter uma taxa de reconhecimento de aproximadamente 70% utilizando uma câmera de alta definição com dimensões de 64×64 *pixels*, enquanto que utilizando uma câmera de vigilância, na mesma distância, foi possível obter uma taxa de reconhecimento de aproximadamente 11%, com a técnica SRC, utilizando dimensões de 16×16 *pixels*.

O trabalho proposto pelos autores se relaciona com esta pesquisa em diversos fatores, que serão melhor apresentados e discutidos durante o decorrer deste trabalho. Cabe, no entanto, citar previamente alguns aspectos relacionados como, por exemplo, os diferentes tipos de câmeras de segurança, a distância entre o sujeito e a câmera, bem como a qualidade das imagens obtidas, fatores estes que influenciam diretamente no desempenho do reconhecimento facial.

A principal diferença que pode ser apontada entre esta pesquisa e o trabalho apresentado pelos autores está no ambiente e nas câmeras utilizadas, pois enquanto no trabalho apresentado os testes realizados são feitos com a câmera situada em um corredor onde os sujeitos se encontravam geralmente de frente para ela, com movimentação relativamente lenta e constante, nesta pesquisa, a câmera está alocada em um canto da sala, muitas vezes sem ter o foco na face dos alunos e, como eles estarão em constante movimentação, possivelmente ocorrerão problemas de oclusão.

2.3 Considerações

O Quadro 1 apresenta uma comparação entre algumas variáveis consideradas e técnicas utilizadas em relação aos trabalhos relacionados e o trabalho aqui descrito.

Considere a seguinte legenda para o Quadro 1:

- **Trabalho 1:** *Who missed the class?-Unifying multi-face detection, tracking and recognition in videos* (MAO; LI; YIN, 2014).
- **Trabalho 2:** *Pose-robust recognition of low-resolution face images* (BISWAS et al., 2013).

- **Trabalho 3:** *RedFace*: um sistema de reconhecimento facial baseado em técnicas de análise de componentes principais e autofaces (DINIZ et al., 2013).
- **Trabalho 4:** *Low resolution face recognition in surveillance systems* (XU; LIU; LI, 2014).
- **O Presente Trabalho:** Comparação de Técnicas de Reconhecimento Facial para Identificação de Presença em um Ambiente Real e Semicontrolado.

Quadro 1 – Comparação das variáveis consideradas e das técnicas utilizadas

	Trabalho 1	Trabalho 2	Trabalho 3	Trabalho 4	O Presente Trabalho
Posição da Câmera	Móvel	Fixa	Fixa	Fixa	Fixa
Resolução das Imagens de Teste	-	Baixa	-	Variada	Baixa
Resolução das Imagens de Treinamento	-	Alta	-	Variada	Variada
Variação de Poses	Sim	Sim	Não	Sim	Sim
Variação de Iluminação	Sim	Sim	Sim	Sim	Sim
Rastreamento	Sim	Sim	Não	Não	Não
Técnicas Utilizadas	Eigenfaces, LBP, K-Means	Tensor Analysis, SIFT, PCA	Eigenfaces, Random Forest, K-NN, K-Star	LDA, LPP, PCA, SRC	Eigenfaces, Fisherfaces, LBPH, SIFT, SURF
Bases de Dados Utilizadas	Honda/UCSD, Base Própria	Multi-PIE, ScFace, MBGC	Base Própria	FRGC, ScFace, AR, Curtin Faces	Base Própria

Fonte: Kelvin Salton do Prado, 2017.

Por meio das informações apresentadas no Quadro 1, é possível concluir que o presente trabalho diferencia-se dos trabalhos relacionados pelo fato de considerar um conjunto de variáveis não tratadas em sua totalidade, bem como pela utilização e comparação de diferentes técnicas de reconhecimento facial. Como já mencionado nas seções anteriores, tais variáveis referem-se principalmente à baixa qualidade das imagens de teste, variação de iluminação, movimentação constante e oclusão. Estes pontos se mostram como grandes desafios para o reconhecimento facial e prejudicam significativamente a precisão do reconhecimento. Acredita-se que os resultados obtidos com o presente trabalho poderão auxiliar futuros trabalhos nesta área, principalmente aqueles realizados em ambientes reais, onde não se tem um controle total sobre o ambiente e seus frequentadores.

3 Fundamentação Teórica

Neste capítulo será apresentada a fundamentação teórica correspondente à detecção e ao reconhecimento facial, bem como os fundamentos básicos sobre o funcionamento dos algoritmos utilizados neste trabalho.

3.1 *Detecção facial*

Antes de realizar o reconhecimento facial em uma determinada imagem é necessário localizar (ou detectar) e extrair as imagens faciais contidas na própria imagem, processo este popularmente conhecido como detecção facial (SZELISKI, 2010).

As técnicas de detecção facial podem ser classificadas como: baseada em características, baseada em modelo e baseada em aparência (YANG; KRIEGMAN; AHUJA, 2002). A técnica baseada em características tenta encontrar a localização de características distintas da imagem, como, por exemplo, os olhos, o nariz e a boca, e então verifica se estas características estão em um arranjo geometricamente plausível, concluindo se aquele conjunto de características representa ou não uma face (SZELISKI, 2010).

A abordagem baseada em modelo, por sua vez, pode lidar com uma ampla gama de variação de pose e expressão. Normalmente, esta abordagem requer uma boa inicialização próxima de uma face real e, portanto, não é recomendada ou adequada como um detector facial rápido (SZELISKI, 2010). Por fim, a abordagem baseada em aparência digitaliza pequenas manchas retangulares na imagem, procurando por prováveis candidatos a faces, que então podem ser refinados utilizando um algoritmo de detecção de modelo em cascata, computacionalmente mais pesado, porém mais seletivo. Nos tempos atuais, a maior parte das abordagens baseadas em aparência dependem fortemente de classificadores de treinamento usando conjuntos de padrões de imagens faciais e não faciais rotuladas. Um exemplo de algoritmo que utiliza a abordagem baseada em aparência é o algoritmo de Viola-Jones (SZELISKI, 2010).

Para a realização da detecção facial neste trabalho foi aplicado o método Viola-Jones, apresentado na próxima seção, utilizando um algoritmo de detecção facial de modelo em cascata disponibilizado pela biblioteca OpenCV (Open Source Computer Vision). Este

método foi selecionado por ser bem difundido na literatura (PRADO et al., 2016) e por ser disponibilizado pela biblioteca OpenCV.

3.1.1 Algoritmo Viola-Jones

A técnica de Viola-Jones é uma abordagem de aprendizado de máquina para a detecção de objetos ou, neste caso, faces em imagens, de forma rápida e com alta taxa de precisão (VIOLA; JONES, 2001). A abordagem de Viola-Jones propõe o uso das chamadas “características de Haar” para realizar a classificação das imagens. Mais precisamente, os autores utilizam três tipos de características, sendo elas:

- Dois retângulos: o valor da função é a diferença entre a soma da intensidade dos *pixels* dentro das duas regiões dos retângulos. As regiões têm o mesmo tamanho e forma e são horizontalmente ou verticalmente adjacentes (Figuras 3 A e 3 B).
- Três retângulos: calcula a soma dos dois retângulos de fora subtraído pela soma do retângulo central (Figura 3 C).
- Quatro retângulos: calcula a diferença entre os pares diagonais de retângulos (Figura 3 D).

Figura 3 – Retângulos de características. Dois retângulos A e B, três retângulos C e quatro retângulos D.

Fonte: Kelvin Salton do Prado, 2016.

As características dos retângulos são computadas rapidamente utilizando uma representação intermediária da imagem, chamada de “imagem integral”. A localização x e y da imagem integral (II) contém a soma dos *pixels* acima e à esquerda da posição x e y da imagem original (I), além do próprio *pixel* (x, y) , ou seja:

$$II_{(x,y)} = \sum_{i=0}^{x} \sum_{j=0}^{y} I_{(i,j)}, \quad (1)$$

onde $II(x, y)$ representa um *pixel* da imagem integral e $I(x, y)$ representa um *pixel* da imagem original. Um exemplo desta operação pode ser visualizado na Figura 4, onde cada valor da imagem integral corresponde à soma da intensidade dos *pixels* à esquerda e acima da imagem original, assim como pode ser visualizado na posição D da imagem integral, a qual corresponde à soma da intensidade de todos os *pixels* da imagem original ($10 + 15 + 18 + 20$).

Figura 4 – Exemplo do cálculo da imagem integral

10	15		
18	20		
Imagem Original			

10	25		
A	B		
28	63		
C	D		
Imagem Integral			

Fonte: Kelvin Salton do Prado, 2017.

A principal vantagem desta abordagem é que ela possibilita, com apenas uma varredura na imagem original, realizar a soma dos elementos de qualquer retângulo utilizando apenas quatro pontos (VIOLA; JONES, 2001; ARAUJO, 2010; COSTA, 2016). Levando em consideração a imagem integral gerada na Figura 4, é possível calcular os elementos do retângulo utilizando a seguinte equação:

$$I(D) = D - B - C + A \quad (2)$$

Como a detecção facial trata-se de um problema relativamente complexo, a utilização de um único classificador não é capaz de alcançar o desempenho esperado e, por este motivo, utiliza-se um classificador em cascata. O algoritmo Viola-Jones emprega uma variação do método *AdaBoost*, responsável por selecionar um conjunto de características para treinar o classificador e, por meio de um conjunto de classificadores fracos, construir um sistema com alta taxa de confiabilidade (VIOLA; JONES, 2001; COSTA, 2016).

O conceito básico do procedimento em cascata para alcançar uma alta taxa de confiabilidade é de que a avaliação de um segundo classificador só será invocada se a avaliação do primeiro classificador foi positiva, caso contrário, o processo é interrompido. Assim, se faz necessário obter um resultado positivo em todos os classificadores para que a

detecção do padrão de uma janela tenha sucesso (VIOLA; JONES, 2004 apud SANTANA; ROCHA, 2015)(COSTA, 2016). Este processo é apresentado na Figura 5, onde é utilizada uma janela da imagem como entrada para o primeiro classificador (C1) e, apenas se ela for considerada verdadeira (imagem facial), ela será submetida ao próximo classificador (C2) e assim sucessivamente até o último classificador (Cn). Sendo assim, se ao menos um dos classificadores categorizar a janela como falsa (imagem não facial) a mesma será desconsiderada.

Figura 5 – Esquema do classificador em cascata

Fonte: Kelvin Salton do Prado, 2017.

Apesar dos bons resultados obtidos pelo uso do método Viola-Jones e do avanço tecnológico dos sistemas de biometria facial em geral, algumas dificuldades ainda são enfrentadas como, por exemplo, a dificuldade da localização de faces em ambientes não controlados (DHARAVATH; TALUKDAR; LASKAR, 2013 apud COSTA, 2016).

O algoritmo Viola-Jones foi utilizado neste trabalho por meio da biblioteca OpenCV, a qual disponibiliza um conjunto de classificadores “pré-treinados” através de arquivos XML que podem ser importados para o algoritmo, não sendo necessário realizar todo o processo de treinamento manualmente. Neste contexto, em relação a este trabalho, acredita-se não ser viável a criação de um conjunto de treinamento próprio para o algoritmo de detecção facial, pois seriam necessárias milhares de imagens, incluindo imagens faciais e não faciais, ou seja, imagens diversas, para se obter um resultado relevante que pudesse ser utilizado no trabalho. Um exemplo disto é apresentado no trabalho de Viola e Jones (2004), no qual os autores utilizam 15.000 imagens para treinamento do algoritmo de detecção, sendo 5.000 imagens faciais e 10.000 imagens não faciais.

3.2 Reconhecimento facial

Dentre as mais diversas tarefas de reconhecimento executadas pelos computadores, o reconhecimento facial é uma das que mais demonstrou sucesso até os dias atuais. Mesmo que os computadores não possam distinguir sujeitos entre milhares de pessoas em uma sequência de vídeo, sua capacidade de distinguir indivíduos em um pequeno grupo tem evoluído bastante e encontrado diversas aplicações de nível comercial. Além disto, o reconhecimento facial pode ser utilizado em uma variedade enorme de aplicações, incluindo a interação humano computador¹, a verificação de identidade, o monitoramento de pacientes, entre outras áreas (SZELISKI, 2010).

Os métodos de reconhecimento facial atuais trabalham melhor utilizando imagens faciais frontais e em condições de iluminação uniformes. Entretanto, bases de imagens reais coletadas podem incluir grandes variações de pose e de iluminação (SZELISKI, 2010).

Algumas das primeiras abordagens de reconhecimento facial conhecidas consistiam em localizar características distintas na imagem como, por exemplo, os olhos, nariz e boca, e então medir as distâncias entre estes pontos de características. Porém, algumas das abordagens de reconhecimento facial mais recentes trabalham na comparação de imagens em escala de cinza, projetadas em subespaços de baixa dimensão conhecidos como *eigenfaces* (SZELISKI, 2010). Nas seções seguintes serão descritos os métodos de reconhecimento facial utilizados neste trabalho.

3.2.1 Eigenfaces

O algoritmo Eigenfaces, proposto por Turk e Pentland (1991), é um dos algoritmos mais difundidos na área de reconhecimento facial. Basicamente, o algoritmo projeta imagens faciais em um espaço de características que abrange as variações significativas entre imagens faciais já conhecidas. As características são conhecidas como *eigenfaces*, pois elas são os autovetores das componentes principais do conjunto de faces.

A abordagem do Eigenfaces busca extrair informações relevantes de uma imagem facial, codificando isso da forma mais eficiente possível, e comparando a similaridade de uma face codificada com uma base de modelos codificados. Matematicamente, a abordagem do Eigenfaces busca encontrar as componentes principais da distribuição das faces ou os

¹ Do inglês *Human-Computer Interaction* (HCI).

autovetores da matriz de covariância do conjunto de imagens, tratando a imagem como um vetor.

Calculando as Eigenfaces

Conforme descrito por Turk e Pentland (1991), o cálculo das *eigenfaces* parte do princípio de que se tem uma imagem facial bidimensional de tamanho N por N *pixels* em escala de cinza. Esta imagem pode então ser representada como um vetor N^2 . Assim, uma imagem de dimensões 256 de largura por 256 de altura, por exemplo, poderia ser representada como um vetor linha de 65.536 posições. Com o propósito de diminuir a dimensionalidade dos dados, é utilizada a técnica de Análise de Componentes Principais², transformando os dados em um subespaço de relativamente baixa dimensionalidade. A ideia básica da análise de componentes principais neste cenário é encontrar os vetores que melhor representam a distribuição das imagens faciais dentro de todo o espaço da imagem.

Durante o processo de treinamento, a técnica de *eigenfaces* calcula uma “face média”, ou seja, um vetor linha μ com a média de todas as faces utilizadas no treinamento. Supondo que M é o número de imagens faciais para treinamento e Γ é o vetor linha que representa cada imagem facial, tem-se então a seguinte fórmula:

$$\mu = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (3)$$

O resultado será um vetor médio μ , o qual representa a “face média” das imagens utilizadas no treinamento. A Figura 6 representa um exemplo da face média (direita), gerada a partir do resultado de μ que, por sua vez, foi calculado a partir das imagens da base de treinamento (esquerda). Cabe esclarecer que, como as imagens são em escala de cinza, os valores utilizados nos cálculos representam a intensidade de cada *pixel*.

² Do inglês *Principal Component Analysis* (PCA)

Figura 6 – Face média calculada a partir da base de treinamento.

Fonte: Kelvin Salton do Prado, 2017.

Após esta etapa, para cada vetor de imagem é calculado um novo vetor (Φ), subtraindo o vetor médio (μ), calculado na etapa anterior, do vetor de cada imagem facial (Γ), como pode-se observar na Fórmula 4.

$$\Phi_i = \Gamma_i - \mu \quad (4)$$

Constrói-se então a matriz A , onde cada coluna i da matriz representa o vetor linha Φ_i , tal que $A = [\Phi_1 \Phi_2 \dots \Phi_M]$. Em seguida, é calculada a matriz de covariância 5:

$$C = \frac{1}{M} \sum_{i=1}^M \Phi_i \Phi_i^T \quad (5)$$

$$C = AA^T$$

Para encontrar as *eigenfaces* é preciso calcular os autovetores e os autovalores da matriz C . Porém, a matriz C ainda é relativamente grande ($N^2 \times N^2$), o que torna a tarefa inviável devido ao alto custo computacional.

De acordo com Turk e Pentland (1991), caso o número de imagens seja menor que a dimensão do espaço ($M < N^2$), haverá apenas $M-1$, em vez de N^2 , vetores significativos. Os autovetores restantes terão autovalores zero associados. Assim, é possível resolver para autovetores de dimensão N^2 , neste caso, primeiramente resolvendo os autovetores para a matriz $M \times M$ (e.g. resolvendo uma matriz de dimensões 16×16 em vez de uma matriz de dimensões 16.384×16.384), e então utilizar combinações lineares adequadas das imagens

Φ_i . Neste contexto, para construir a matriz $M \times M$ é possível utilizar a seguinte análise, considerando os autovetores v_i de AA^T , tais que:

$$A^T Av_i = \lambda_i v_i \quad (6)$$

Pré-multiplicando ambos os lados por A , tem-se:

$$AA^T Av_i = \lambda_i Av_i \quad (7)$$

Pode-se observar que Av são os autovetores de $C = AA^T$.

Seguindo esta análise, é possível construir a matriz $M \times M$, $L = A^T A$, onde $L_{mn} = \Phi_m^T \Phi_n$, e encontrar os M autovetores, v_l , de L . Estes vetores determinam combinações lineares do conjunto de imagens faciais de treinamento (M) para formar as *eigenfaces* u_l :

$$u_l = \sum_{i=1}^M v_{li} \Phi_i, l = 1, \dots, M \quad (8)$$

Neste caso, u_l representa as *eigenfaces* calculadas, para cada imagem facial de treinamento (M), a partir do cálculo dos autovetores v_l multiplicados pelos vetores de cada imagem (Φ). Desta forma os cálculos são verdadeiramente reduzidos da ordem do número de *pixels* na imagem (N^2) para a ordem do número de imagens no conjunto de treinamento (M). Geralmente, o conjunto de imagens faciais de treinamento é relativamente pequeno, se comparado com o número de *pixels* nas imagens, o que torna os cálculos mais viáveis. Os autovalores associados nos permitem classificar os autovetores de acordo com a sua utilidade em caracterizar a variação entre as imagens (TURK; PENTLAND, 1991).

As imagens *eigenfaces*, calculadas através dos autovetores de L , abrangem um conjunto de base com o qual é possível descrever as imagens faciais. Sendo assim, é possível realizar o reconhecimento facial. Na prática, um número pequeno de imagens (M) é suficiente para realizar o reconhecimento, não sendo necessário realizar a reconstrução precisa da imagem. Desta forma, o reconhecimento facial se torna uma tarefa de reconhecimento de padrões, onde as *eigenfaces* abrangem um subespaço dimensional (M) do espaço original (N). Os M' autovetores significativos da matriz L são escolhidos como aqueles com os maiores autovalores associados (TURK; PENTLAND, 1991).

Considerando $i = 1, \dots, M'$, a nova imagem facial (Γ) é transformada em seus componentes *eigenfaces* pela seguinte operação:

$$\omega_i = u_i^T (\Gamma - \mu) \quad (9)$$

Os pesos (ω_i) são então calculados, utilizando-se cada uma das *eigenfaces* (u_i^T), o vetor da imagem (Γ) e o vetor médio (μ). Segundo Turk e Pentland (1991), os pesos (ω_i) formam o vetor $\Omega^T = [\omega_1, \omega_2, \dots, \omega_m]$, que descreve a contribuição de cada *eigenface* em representar a imagem facial de entrada, tratando as *eigenfaces* como um conjunto base para as imagens faciais. O vetor pode então ser utilizado em um algoritmo de reconhecimento de padrões para descobrir, baseado em um número de classes pré-definido, qual descreve melhor a face. O método mais simples para determinar qual classe fornece a melhor descrição de uma imagem facial de entrada é encontrar a classe k que minimiza a distância Euclidiana entre os vetores, o que pode ser feito por meio da função:

$$\epsilon_k = ||(\Omega - \Omega_k)||^2, \quad (10)$$

onde Ω_k é um vetor que descreve a classe. As classes Ω_i são calculadas pela média dos resultados da representação *eigenfaces* sobre um pequeno número de imagens faciais de cada indivíduo. A face é classificada como pertencente a classe k quando o ϵ_k mínimo é inferior a um limiar definido. Caso contrário, a face é classificada como “desconhecida” (TURK; PENTLAND, 1991).

3.2.2 Local Binary Patterns Histograms (LBPH)

Originalmente o *Local Binary Patterns* (LBP), introduzido por Ojala, Pietikäinen e Harwood (1996), foi desenvolvido como um método para análise de texturas. Porém, com o tempo, passou a ser utilizado para a extração de características no processo de reconhecimento e classificação de imagens faciais. Basicamente, o LBP rotula os *pixels* de uma imagem levando em consideração uma matriz, geralmente de tamanho 3×3 , sendo que o *pixel* central da matriz é utilizado como limiar para definir o valor dos *pixels* vizinhos, os quais são definidos de forma binária (AHONEN; HADID; PIETIKÄINEN, 2004; AMARAL; GIRALDI; THOMAZ, 2013).

Considerando parte de uma imagem como uma matriz de tamanho 3×3 *pixels* e em escala de cinza (Figura 7a), a qual pode ser representada como uma matriz de inteiros que representam a intensidade do *pixel* (Figura 7b), também de tamanho 3×3 , o LBP rotula cada *pixel* de acordo com o valor central da matriz. Este é então utilizado como limiar para definir o valor dos *pixels* vizinhos, que por sua vez são rotulados de forma binária, sendo atribuído o valor 1 para os *pixels* que possuem um valor igual ou superior ao limiar e 0 para os *pixels* que possuem valor inferior ao limiar (Figura 7c) (AHONEN; HADID; PIETIKÄINEN, 2004). Os valores binários resultantes são então concatenados formando um novo valor binário, valor este que é então convertido para o sistema decimal e atribuído à posição central da matriz (Figura 7d), sendo que esta representa parte da imagem em escala de cinza.

Figura 7 – Exemplo de operação básica do LBP.

Fonte: Kelvin Salton do Prado, 2016.

A Figura 8 apresenta um exemplo do resultado desta operação, sendo que a imagem à esquerda corresponde à imagem original, utilizada como entrada no algoritmo, enquanto que a imagem à direita corresponde ao resultado da operação.

Figura 8 – Exemplo de imagem resultante da operação básica do LBP.

Fonte: Kelvin Salton do Prado, 2016.

A nova imagem gerada (Figura 9b) é então dividida em regiões, geralmente organizadas em uma malha regular (Figura 9c), e os padrões (também chamados de descritores de texturas por alguns autores) são extraídos de cada região isoladamente (Figura 9d). Ao final, os descritores são concatenados para formar um descritor global da face (Figura 9e), chamado de histograma da imagem (AHONEN; HADID; PIETIKAINEN, 2006), conforme ilustra a Figura 9.

Figura 9 – Extração do histograma da face.

Fonte: Kelvin Salton do Prado, 2016.

Após extrair o histograma de cada imagem facial, a classificação é então realizada através do cálculo de similaridade entre os histogramas. Cabe esclarecer que, além dos histogramas serem utilizados como descritores, eles também auxiliam na diminuição da dimensionalidade dos dados. Neste contexto, é possível representar os padrões de uma imagem de qualquer dimensão utilizando um vetor relativamente menor. Por exemplo, supondo que se tenha uma imagem de dimensões 300×300 , o que corresponde a um total de 90.000 *pixels*, é possível representá-la, considerando uma malha regular de tamanho 8×8 , utilizando um vetor de 16.384 posições, o qual é composto por 64 histogramas de 256 posições cada ($8 \times 8 \times 256 = 16.384$). Este processo é conhecido também como *Local Binary Patterns Histograms* (LBPH).

Posteriormente, os operadores foram estendidos para utilizar vizinhanças de diferentes tamanhos. Utilizando vizinhanças circulares e interpolação bilinear, os valores dos *pixels* permitem qualquer raio e números de pontos na vizinhança (AHONEN; HADID; PIETIKÄINEN, 2004). A notação mais comum utilizada é a letra P ($P > 1$) para a quantidade de *pixels* vizinhos, que significa o número de pontos na amostra, e R ($R > 0$) que representa o raio do círculo (AHONEN; HADID; PIETIKAINEN, 2006). A Figura 10 exemplifica o uso do LBP circular, apresentando exemplos de conjuntos de vizinhanças circulares para

diversos pontos (P) e raios (R). Cabe esclarecer que, quando algum ponto (P) não fica exatamente no centro de um pixel, o valor deste ponto é estimado por interpolação (OJALA; PIETIKAINEN; MAENPAA, 2002).

Figura 10 – Exemplo de LBP circular.

Fonte: Kelvin Salton do Prado, 2016.

Neste contexto, de acordo com Ojala, Pietikainen e Maenpaa (2002), a textura T é definida, com base na vizinhança local, como a distribuição conjunta dos níveis de cinza dos P pixels da imagem:

$$T = t(g_c, g_0, \dots, g_{P-1}), \quad (11)$$

onde g_c corresponde ao ponto central da vizinhança local e $g_p (p = 0, \dots, P-1)$ correspondem aos P valores da vizinhança circular de raio R , como pode ser visualizado na Figura 10 (a). Ainda, caso as coordenadas de g_c sejam $(0, 0)$, as coordenadas de g_p são dadas por $(-R \sin(2\pi p/P), R \cos(2\pi p/P))$.

Com base na descrição apresentada anteriormente, acredita-se que a utilização de um número maior de vizinhança (P) irá resultar em um número maior de padrões obtidos, porém, consequentemente, irá aumentar o custo computacional. Por outro lado, utilizando um raio maior (R), maiores detalhes podem ser capturados, porém o descritor será menos discriminativo, principalmente se não for aumentado o número da vizinhança proporcionalmente. Neste cenário cabe ainda esclarecer que o método LBPH disponibilizado pela biblioteca OpenCV utiliza como padrão um número de vizinhança igual a 8 e um raio igual a 1.

O método LBPH foi selecionado por ser um algoritmo conhecido e utilizado de acordo com a literatura pesquisada (e.g. (AHONEN; HADID; PIETIKAINEN, 2006; LOUIS; PLATANOTIS, 2010; LOUIS; PLATANOTIS; RO, 2010)) e por ser disponibilizado pela biblioteca OpenCV.

3.2.3 Fisherfaces

O método conhecido como *fisherfaces*, proposto por Belhumeur, Hespanha e Kriegman (1997), é baseado na Análise de Discriminante Linear (LDA) e foi desenvolvido com a ideia de ser um método não sensível à grande variação de iluminação e expressões faciais. Assim como o *eigenfaces*, o *fisherfaces* também se baseia na redução de dimensionalidade no espaço de características (SOUZA, 2014). Resultados de experimentos realizados por Belhumeur, Hespanha e Kriegman (1997) demonstraram que o método *fisherfaces* teve taxas de erro menores que o *eigenfaces*, em relação aos testes com as bases de dados faciais de Harvard e de Yale. Apesar de o método ser insensitivo à grande variação de iluminação e expressões faciais, ele não tem a proposta de lidar com variações de pose, o que pode ser um fator crucial no reconhecimento facial em ambientes reais.

Com um conjunto de treinamento rotulado, faz sentido utilizar esta informação para construir um método mais confiável para reduzir a dimensionalidade do espaço de características (BELHUMEUR; HESPANHA; KRIEGMAN, 1997). Na Análise de Discriminante Linear, o conjunto de imagens de treinamento é formado por diversas classes, as quais representam a identidade de cada sujeito com base em um conjunto previamente conhecido. Dada uma nova imagem, o problema de reconhecimento facial é elaborado de modo a determinar a que classe pertence a imagem fornecida (SOUZA, 2014).

No âmbito do reconhecimento facial, a LDA pode ser entendida como uma técnica que busca maximizar a variância entre as classes e minimizar a variância dentro das classes, garantindo uma maior separabilidade entre as classes de um conjunto de imagens, sendo aplicada uma transformação linear visando a encontrar um sistema de coordenadas que melhor represente as imagens (SANTOS, 2005 apud SOUZA, 2014).

Primeiramente, o método *fisherfaces* projeta o conjunto de imagens em um espaço de baixa dimensionalidade utilizando a técnica de PCA, aplicando então a técnica de LDA para projetar as imagens faciais em um espaço de dimensionalidade ainda menor. Por meio disto, o reconhecimento facial é realizado com a projeção no subespaço dos vetores próprios, utilizando uma medida de semelhança para comparar com as demais faces projetadas no subespaço (FIGUEREDO, 2011 apud SOUZA, 2014).

Supondo que Z_{ij} seja uma imagem facial representada por uma matriz de tamanho M , contendo os valores da intensidade dos *pixels* correspondentes a imagem, projetada no subespaço PCA, a média total (μ) das imagens faciais é dada por (SOUZA, 2014):

$$\mu = \frac{1}{\sum_{i=1}^r N_i} \sum_{i=1}^r \sum_{j=1}^{N_i} Z_{ij}, \quad (12)$$

onde N_i é a quantidade de imagens faciais pertencentes a cada uma das r classes. Ainda, com o intuito de facilitar a compreensão, é possível simplificar a fórmula da seguinte maneira:

$$\mu = \frac{1}{N} \sum_{i=1}^N Z_i, \quad (13)$$

onde N corresponde ao número total de imagens da base de treinamento. Além disto, a face média (μ_i) de cada classe é obtida através da expressão (SOUZA, 2014):

$$\mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Z_{ij} \quad (14)$$

A matriz de dispersão interclasse S_B é, por sua vez, definida por:

$$S_B = \sum_{i=1}^r N_i (\mu_i - \mu)(\mu_i - \mu)^T, \quad (15)$$

a qual representa a dispersão das faces médias de cada classe em relação a média total (SOUZA, 2014).

A matriz de dispersão intraclasse S_w é definida por:

$$S_W = \sum_{i=1}^r \sum_{j=1}^{C_i} (Z_{ij} - \mu_i)(Z_{ij} - \mu_i)^T = \sum_{i=1}^r S_{W_i}, \quad (16)$$

onde S_{W_i} é uma matriz de covariância de uma determinada classe (C_i) e S_W representa a dispersão das imagens de faces Z_{ij} em relação à face média de sua classe.

Com base em Kami (2011), Souza (2014) esclarece que:

A técnica LDA busca encontrar uma matriz de projeção ótima que maximize a razão do determinante da matriz de dispersão interclasse das imagens de faces projetadas e também que minimize o determinante da matriz de dispersão intraclasse dessas mesmas imagens (KAMI, 2011 apud SOUZA, 2014).

A matriz de projeção ótima (W_{otima}) é definida por:

$$W_{otima} = arg \max_W \frac{|W^T \cdot S_B \cdot W|}{|W^T \cdot S_W \cdot W|} = [W_1 \ W_2 \ \dots \ W_M], \quad (17)$$

onde $\{W_i | i = 1, 2, \dots, M\}$ é o conjunto generalizado de autovetores de S_B e S_W correspondente aos M maiores autovalores $\{\lambda_i | i = 1, 2, \dots, M\}$ (BELHUMEUR; HESPANHA; KRIEGMAN, 1997; SOUZA, 2014).

Resumidamente, o método *fisherfaces* utiliza PCA para reduzir a dimensionalidade, projetando as imagens faciais originais em um espaço de menor dimensão e, em seguida, utiliza a técnica LDA para procurar a melhor característica discriminante linear no subespaço de PCA (SILVA et al., 2006 apud SOUZA, 2014).

Formalmente, a matriz de projeção ótima (W_{otima}) pode ser calculada da seguinte forma:

$$W_{otima} = W_{LDA} W_{PCA}, \quad (18)$$

onde

$$W_{PCA} = arg \max_W |W^T \cdot S_{BW} \cdot W|, \quad (19)$$

para $S_{BW} = S_B + S_W$ e W_{LDA} é dada por:

$$W_{LDA} = arg \max_W \frac{|W^T \cdot W_{PCA}^T \cdot S_B \cdot W_{PCA} \cdot W|}{|W^T \cdot W_{PCA}^T \cdot S_W \cdot W_{PCA} \cdot W|} \quad (20)$$

Assim sendo, é possível realizar a classificação de uma nova imagem facial comparando a similaridade das novas imagens de teste com as imagens faciais do conjunto de treinamento, determinando assim a qual sujeito a imagem facial pertence.

Belhumeur, Hespanha e Kriegman (1997) argumentam que, com o uso de métodos lineares específicos de classe para redução de dimensionalidade, além de classificadores simples no espaço de características reduzido, pode-se obter melhores taxas de reconhecimento se comparados a outros métodos como *Linear Subspace* e *Eigenfaces*. Ao final, os autores concluem que o método *Fisherfaces* obteve melhores resultados nos testes realizados, utilizando imagens com variação de iluminação e expressão facial, com relação aos métodos *Eigenfaces* e *Linear Subspace* (BELHUMEUR; HESPANHA; KRIEGMAN, 1997). O método *Fisherfaces* foi utilizado neste trabalho por ser um método conhecido na literatura (e.g. (TANG; LYU; KING, 2003; YANG, 2006)) e por ser disponibilizado pela biblioteca OpenCV.

3.2.4 Scale Invariant Feature Transform (SIFT)

O método *Scale Invariant Feature Transform* (SIFT) foi publicado originalmente por Lowe (1999) como um método para extração de características invariantes em imagens, que posteriormente podem ser utilizadas para realizar uma correspondência³ confiável entre diferentes visões de um objeto ou de uma cena (LOWE, 2004; LOWE, 1999). De acordo com Lowe (1999) as características extraídas são invariantes à escala e à rotação da imagem, além de parcialmente invariantes a alterações na iluminação e no ponto de vista da câmera.

Apesar de inicialmente o SIFT ter apresentado uma abordagem utilizando a extração de características para a realização do reconhecimento de objetos em uma cena, posteriormente o método passou a ser utilizado em diversas aplicações como, por exemplo, mapeamento em robótica para navegação, modelagem 3D, reconhecimento de gestos, reconhecimento facial, entre outras (GONZÁLES, 2010).

Segundo Lowe (2004) é possível extrair muitas características de uma imagem utilizando algoritmos eficientes, uma vez que as características são altamente distintas, no sentido de que uma única característica pode ser combinada corretamente, com uma alta probabilidade, contra uma base de dados de características, tornando assim possível o reconhecimento de objetos em uma cena. O custo de extração das características é então minimizado por meio de uma abordagem de filtragem em cascata, na qual as operações mais dispendiosas são aplicadas apenas em locais que passam por um teste inicial (LOWE, 2004).

Dentro do SIFT, as principais etapas computacionais utilizadas para gerar o conjunto de características da imagem são:

1. Detecção do Espaço-Escala: A detecção do espaço-escala⁴ é a primeira etapa computacional e tem como objetivo pesquisar por pontos de interesse em todas as escalas e locais da imagem. Este processo é eficientemente implementado utilizando a função de Diferença de Gaussianas⁵ para identificar potenciais pontos de interesse que são invariantes à escala e orientação (LOWE, 2004).

³ Do inglês *matching*.

⁴ Do inglês *scale-space extrema detection*.

⁵ Do inglês *Difference of Gaussian* (DoG).

O espaço-escala de uma imagem é definido pela função $L(x, y, \sigma)$, que é produzida pela convolução (indicada pelo operador $*$) de uma função Gaussiana $G(x, y, \sigma)$ com uma imagem de entrada $I(x, y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (21)$$

onde a função Gaussiana é representada por:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (22)$$

Para detectar eficientemente a localização dos pontos de interesse estáveis em um espaço-escala, Lowe (2004) propôs a utilização da função de Diferença de Gaussianas (DoG), que pode ser computada pela diferença de duas escalas próximas separadas por um fator multiplicativo constante k :

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (23)$$

A função DoG suaviza as imagens e pode ser calculada pela subtração das imagens por um filtro gaussiano em escalas σ e $k\sigma$. A utilização da função gaussiana tem como objetivo obter amostras da imagem, onde detalhes indesejados e ruídos são eliminados e as características mais significantes são mantidas, sendo possível encontrar tais características em diferentes escalas por meio da variação do σ (LOWE, 2004; GONZÁLES, 2010).

De acordo com Gonzáles (2010), uma abordagem eficiente para a criação da Diferença de Gaussianas (DoG) é apresentada na Figura 11. Neste contexto, deseja-se construir s intervalos, onde cada intervalo representa uma imagem filtrada por DoG (BELO, 2006). Ainda, para se construir s intervalos, o fator k é definido por $k = 2^{1/s}$, produzindo assim $s + 3$ imagens da oitava, de forma que a detecção de extremos cubra toda a oitava (BELO, 2006; GONZÁLES, 2010). Sendo assim, a imagem inicial sofre convoluções incrementais com filtros Gaussianos para produzir imagens separadas por um fator de escala k no espaço de escala. Gonzáles (2010) explica que Lowe (2004) considera necessário fazer a convolução da imagem até 2σ para ser possível a construção de descritores invariantes em relação à escala. Cabe esclarecer que, de acordo com Belo (2006), cada oitava representa um conjunto de imagens para a

imagem reescalada com diferentes amostragens, como pode ser visto no exemplo da Figura 11. Imagens em escalas adjacentes são subtraídas para produzir as imagens da Diferença das Gaussianas, apresentadas à direita na Figura 11. Uma vez processada a oitava, é reduzida a dimensão da imagem, tomando-se cada segundo *pixel* da imagem no centro da oitava, gerando-se uma nova oitava e voltando-se ao passo 1.

Figura 11 – Procedimento de obtenção das Diferenças de Gaussianas (DoG) para diversas oitavas.

Fonte: Gonzáles (2010).

Com isso, é feita a detecção de extremos em cada intervalo de cada oitava. Um extremo é definido como qualquer valor no DoG maior ou menor do que todos os seus vizinhos no espaço-escala (GONZÁLES, 2010).

2. **Localização Precisa de Pontos Chave:** Em cada candidato local, um modelo detalhado é adequado para determinar a localização e escala. Pontos-chave são selecionados com base em medidas de estabilidade (LOWE, 2004). Com a intenção de localizar os máximos e mínimos locais de $D(x, y, \sigma)$, cada ponto amostrado é comparado com seus 8 vizinhos da imagem atual e com seus 9 vizinhos das escalas superior e inferior, conforme ilustra a Figura 12.

Figura 12 – Detecção de extremos no espaço-escala.

Fonte: Lowe (2004).

Neste caso, o ponto é mantido apenas se ele for maior ou menor do que todos os 26 vizinhos. O custo computacional desta operação é relativamente baixo, pois a maior parte dos pontos amostrados serão eliminados logo nas primeiras verificações (LOWE, 2004).

Após este procedimento, é realizado um ajuste detalhado nos dados da imagem local, determinando a posição e a escala. Nesta etapa, os candidatos a pontos chave que possuem baixo contraste são rejeitados, por serem sensíveis a ruídos ou estarem localizados ao longo das bordas (LOWE, 2004 apud SILVA et al., 2012).

Para corrigir este problema, o algoritmo elimina os pontos chave que possuem uma localização fracamente determinada, mas que possuem uma alta resposta na borda (SILVA et al., 2012).

3. Atribuição de Orientação: Ao atribuir uma orientação consistente a cada ponto-chave, com base em propriedades da imagem local, o descritor de ponto-chave pode ser representado com relação a esta orientação, e assim ser invariante à rotação da imagem. A escala do ponto-chave é utilizada para selecionar a imagem suavizada pela Gaussiana L com a escala mais próxima, de modo que toda a computação seja realizada de forma invariante à escala (LOWE, 2004; SILVA et al., 2012).

Para cada amostra da imagem na escala $L(x, y)$, a magnitude ($m(x, y)$) e a orientação ($\theta(x, y)$) do gradiente são calculadas usando a diferença em *pixels*:

$$\begin{aligned} m(x, y) &= \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \\ \theta(x, y) &= \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \end{aligned} \quad (24)$$

É então formado um histograma de orientação a partir das orientações do gradiente dos pontos amostrados dentro da região ao redor do ponto-chave. O histograma possui 36 regiões que cobrem a gama de orientações de 360 graus (LOWE, 2004; GONZÁLES, 2010).

Os picos do histograma correspondem às direções dominantes do gradiente local. O pico mais alto é detectado e então qualquer outro pico local, que corresponda a pelo menos 80% do pico mais alto, é utilizado também para criar um ponto-chave com esta orientação. Assim sendo, para localizações com múltiplos picos de magnitude similar, haverá múltiplos pontos-chave criados na mesma localização e escala, mas com orientações diferentes.

4. Descritor dos pontos chave: As operações anteriores atribuíram a localização, escala e orientação da imagem para cada ponto-chave. Estes parâmetros impõem um sistema de coordenadas bidimensional (2D) para descrever a região local da imagem e, portanto, fornecer invariância a estes parâmetros (LOWE, 2004). O próximo passo é calcular um descritor para a região da imagem local que é distinto e invariante a mudanças na iluminação ou ponto de vista tridimensional (LOWE, 2004; SILVA et al., 2012).

Para alcançar a invariância de orientação, as orientações do gradiente são rotacionadas em relação à orientação do ponto-chave. O descritor do ponto-chave é então criado calculando-se a magnitude e orientação do gradiente, em cada ponto da imagem amostrada, ao redor da localização do ponto-chave. Este procedimento é apresentado na Figura 13, onde os gradientes são representados pelas pequenas setas (imagem da esquerda) em cada amostra da localização. Os descritores são então ponderados em uma janela de suavização Gaussiana, indicada pelo círculo azul à esquerda da Figura 13. Por fim, estas amostras são acumuladas em histogramas de orientação representado pela imagem da direita da Figura 13.

Figura 13 – Descrição dos pontos chave.

Fonte: Lowe (2004).

Um aspecto importante desta abordagem é que ela gera grandes números de características que cobrem densamente a imagem em toda a gama de escala e localizações. Supõe-se que uma imagem típica, de tamanho 500×500 *pixels*, resulte em cerca de 2.000 características estáveis, embora este número possa variar de acordo com o conteúdo da imagem e as escolhas dos parâmetros. A quantidade de características é particularmente importante para o reconhecimento de objetos, sendo que a habilidade de detectar pequenos objetos em “ambientes desordenados” exige que pelo menos 3 características sejam corretamente correspondidas para uma identificação confiável (LOWE, 2004).

Para o reconhecimento de imagens utilizando o método SIFT, as características são primeiramente extraídas de um conjunto de imagens de referência, geralmente de uma base de treinamento, e armazenadas em uma estrutura de dados. Uma nova imagem é então associada comparando individualmente cada característica da nova imagem com a estrutura de dados criada anteriormente e, assim, encontrando candidatos com características correspondentes com base na distância Euclidiana dos vetores de características (LOWE, 2004).

Os descritores de ponto-chave são altamente distintivos, o que permite que uma única característica encontre sua correspondência correta com uma boa probabilidade em uma grande base de dados de características. No entanto, em uma imagem “desordenada”, muitas características do plano de fundo não terão qualquer correspondência correta com a estrutura de dados, dando origem a muitas correspondências falsas. As correspondências corretas podem ser filtradas do conjunto completo de correspondências, identificando subconjuntos de pontos-chave que concordam com o objeto e sua localização, escala e

orientação na nova imagem. A determinação destes agrupamentos consistentes pode ser realizada rapidamente, utilizando a implementação de uma tabela *hash* da transformada de Hough generalizada (LOWE, 2004).

Cada agrupamento de 3 ou mais características que concordam com um objeto e sua pose são então sujeitos a uma verificação mais detalhada. Primeiramente, uma estimativa dos mínimos quadrados⁶ é feita para uma aproximação da pose do objeto. Quaisquer outras características da imagem consistentes com esta pose são identificadas, e os valores atípicos (*outliers*) são descartados. Finalmente, é feita uma computação detalhada da probabilidade de que um determinado conjunto de características indique a presença de um objeto, dada a precisão do ajuste e o número de prováveis correspondências falsas. Correspondências de objetos que passam por todos estes testes podem ser identificadas como corretas com alta confiança (LOWE, 2004).

3.2.5 Speeded Up Robust Features (SURF)

A detecção de características pode ser definida como o processo responsável por examinar e extrair, de forma automática, características que são únicas para um objeto em uma imagem, geralmente com a pretensão de futuramente realizar a detecção destes objetos em outras imagens e idealmente ser robusto a variações de escala e rotação (PEDERSEN, 2011). A tarefa de encontrar correspondências⁷ entre duas imagens da mesma cena ou objeto é, ainda hoje, parte de diversas aplicações de visão computacional, como por exemplo na reconstrução 3D e reconhecimento de objetos (BAY; TUYTELAARS; GOOL, 2006).

O algoritmo *Speeded Up Robust Features* (SURF), publicado por Bay, Tuytelaars e Gool (2006), foi desenvolvido com base no método SIFT com o objetivo de detectar e descrever pontos de interesse de forma computacionalmente mais rápida, se comparado a outros algoritmos do estado da arte, porém sem perder performance. No algoritmo SURF, a busca por correspondências de pontos de interesse em imagens pode ser dividida basicamente em três etapas principais:

⁶ Do inglês *least-squared*.

⁷ Do inglês *matches*.

- Primeiramente, os “pontos de interesse” são selecionados em locais distintos da imagem, como cantos, *blobs*⁸ e junções em T. A propriedade mais valiosa de um detector de ponto de interesse é sua repetibilidade. A repetibilidade expressa a confiabilidade de um detector para encontrar os mesmos pontos de interesse físicos sob diferentes condições de visualização (BAY; TUYTELAARS; GOOL, 2006).
- Em seguida, a vizinhança de cada ponto de interesse é representada por um vetor de características. O vetor de características, também chamado de descritor, deve ser distinto e ao mesmo tempo robusto a ruídos, erros de detecção e deformações geométricas e fotométricas (BAY; TUYTELAARS; GOOL, 2006).
- Finalmente, os vetores de descritores são combinados entre diferentes imagens. A correspondência é geralmente baseada na distância entre os vetores (por exemplo, a distância Euclidiana ou a distância de Mahalanobis). A dimensão do descritor tem um impacto direto no tempo de execução da tarefa de correspondência, e vetores de menores dimensões são desejáveis para realizar a correspondência de pontos de interesse de forma computacionalmente mais rápida. Contudo, vetores de características de baixa dimensão são em geral menos distintivos que os seus equivalentes de alta dimensão (BAY; TUYTELAARS; GOOL, 2006; BAY et al., 2008).

Detecção de pontos de interesse

O método SURF utiliza uma aproximação de matriz Hessiana básica para realizar a detecção de pontos de interesse. Ele faz uso de imagens integrais, que se tornaram populares por Viola e Jones (2001), para reduzir o tempo de computação (BAY et al., 2008). A entrada de uma imagem integral II em um ponto (x, y) representa a soma da intensidade de todos os *pixels* na imagem de entrada I dentro de uma região retangular formada pela origem e o ponto (x, y) :

$$II_{(x,y)} = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I_{(i,j)} \quad (25)$$

Uma vez que a imagem integral foi calculada, são necessárias apenas 3 adições para calcular a soma das intensidades sobre qualquer área retangular vertical, como pode ser

⁸ *Blobs* podem ser considerados regiões em uma imagem onde algumas propriedades (e.g. cor) são constantes ou aproximadamente constantes e diferem de regiões circundantes.

visto na Figura 14. Desta forma, o tempo de computação é independente do tamanho da região retangular.

Figura 14 – Soma de intensidade dentro de uma região retangular em uma imagem integral.

Fonte: Bay et al. (2008).

Pontos de interesse baseados na matriz Hessiana

SURF utiliza um detector baseado na matriz Hessiana para encontrar os pontos de interesse. O determinante de uma matriz Hessiana expressa a extensão da resposta e é uma expressão da mudança local em torno da área (PEDERSEN, 2011).

Dado um ponto $a = (x, y)$ em uma imagem I , a matriz Hessiana $H(a, \sigma)$ em a na escala σ é definida pela seguinte equação:

$$H(a, \sigma) = \begin{bmatrix} L_{xx}(a, \sigma) & L_{xy}(a, \sigma) \\ L_{xy}(a, \sigma) & L_{yy}(a, \sigma) \end{bmatrix} \quad (26)$$

onde $L_{xx}(a, \sigma)$, $L_{xy}(a, \sigma)$ e $L_{yy}(a, \sigma)$ representam a convolução da imagem com as derivadas de segunda ordem da gaussiana. Os cálculos das convoluções possuem alto custo computacional, e por este motivo é utilizada a imagem integral, que é gerada através da Equação 25 apresentada anteriormente, sendo então possível calcular a área de qualquer retângulo da imagem utilizando apenas 4 variáveis, como exemplificado na Figura 14.

As gaussianas de segunda ordem, utilizadas na matriz hessiana, devem ser discretizadas e recortadas antes de aplicar uma caixa de filtragem 9×9 , como pode ser visto na Figura 15, onde as áreas cinza correspondem a 0, as áreas branca correspondem a um valor positivo e as áreas pretas a um valor negativo (PEDERSEN, 2011).

Figura 15 – Gaussianas discretizadas e recortadas ($L_{yy}(a, \sigma)$ e $L_{xy}(a, \sigma)$) e as aproximações (D_{yy} e D_{xy}).

Fonte: Bay et al. (2008).

Desta forma, é possível calcular a convolução aproximada efetivamente utilizando a imagem integral (BAY et al., 2008; PEDERSEN, 2011):

$$\det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2, \quad (27)$$

onde w representa um peso, que é utilizado para balancear a expressão, necessário para a conservação de energia da gaussiana. O peso w é teoricamente sensível à escala, porém pode-se utilizar a constante 0.9 (BAY et al., 2008; PEDERSEN, 2011).

Para examinar características através da variação de escala, é preciso examinar diferentes oitavas e níveis. Onde o SIFT decrementa a escala da imagem para cada oitava, a imagem integral permite ao SURF calcular respostas com grandes *kernels*. O uso de imagens integrais não só torna isto viável como também torna a operação mais rápida sem a necessidade de reduzir a imagem (PEDERSEN, 2011). Um ponto de interesse é localizado em espaço escala por (x, y, σ) onde x, y são coordenadas relativas e σ é a escala.

Descrição de pontos de interesse

O propósito de um descritor é prover uma descrição única e robusta de uma característica, podendo ser gerado com base na área em torno do ponto de interesse. O descritor do SURF é baseado na *wavelet Haar*, e pode ser calculado eficientemente com imagens integrais. Para isto, é necessário determinar uma única orientação para o ponto

de interesse, sendo assim possível obter invariância rotacional. Antes de ser calculado o descritor, a área de interesse, que encontra-se em torno do ponto de interesse, é rotacionada para esta orientação (PEDERSEN, 2011).

A área de interesse é dividida em 4×4 subáreas, que são descritas pelos valores da Wavelet nas direções x e y , preservando assim importantes informações espaciais. Para cada subárea é computado um vetor v com base em amostras de tamanho 5×5 de cada subárea. O descritor de um ponto de interesse corresponde aos 16 vetores (das 4×4 subáreas) concatenados. Tendo calculados os descritores dos pontos de interesse, encontrar um correspondente torna-se apenas uma questão de calcular a distância entre dois vetores (BAY et al., 2008; PEDERSEN, 2011).

Assim como os demais métodos apresentados, o algoritmo SURF é utilizado neste trabalho através da biblioteca OpenCV.

3.2.6 Brute-Force Matcher

O *Brute-Force Matcher* é um método disponibilizado pela biblioteca OpenCV para fazer a correspondência entre dois conjuntos de descritores de características (OPENCV, 2000). Este método é basicamente utilizado pelos métodos SIFT e SURF para fazer a correspondência entre os descritores da imagem facial que se deseja reconhecer e os descritores das imagens faciais contidas na base de treinamento.

Para um determinado ponto-chave⁹ da primeira imagem, o método compara cada ponto-chave da segunda imagem e calcula a distância entre eles. O ponto-chave da segunda imagem com menor distância será considerado seu par. A medida de distância utilizada para calcular a correspondência pode ser alterada de acordo com um parâmetro disponibilizado pela biblioteca OpenCV. Porém, de acordo com a documentação, recomenda-se utilizar a medida de distância euclidiana (parâmetro chamado de NORM_L2), tanto para o SIFT quanto para o SURF (OPENCV, 2000).

⁹ Do inglês *keypoint*.

3.2.7 Comparação dos algoritmos

No Quadro 2 é apresentada uma breve comparação entre os pontos positivos e negativos dos métodos utilizados neste trabalho, com a intenção de compreender melhor o trabalho como um todo, bem como facilitar o entendimento dos resultados obtidos.

Quadro 2 – Pontos positivos e negativos de cada método

	Ano	Pontos	Descrição
Eigenfaces	1991	Positivos	É um dos métodos mais antigos e conhecidos da área de reconhecimento facial. Insensitivo à pouca variação de iluminação.
		Negativos	A variação de pose e oclusão parcial podem prejudicar significativamente os resultados.
LBPH	1996	Positivos	Robusto em relação a transformações monotônicas em escala de cinza. Permite identificar bordas.
		Negativos	A variação de pose e oclusão parcial podem prejudicar significativamente os resultados.
Fisherfaces	1997	Positivos	Insensitivo a certa variação de iluminação e expressões faciais. Segundo os autores teve melhores resultados do que o eigenfaces.
		Negativos	A variação de pose e oclusão parcial podem prejudicar significativamente os resultados.
SIFT	1999	Positivos	É invariante à escala e rotação, e parcialmente insensitivo à variação de iluminação e oclusão parcial.
		Negativos	É computacionalmente mais pesado que os outros métodos.
SURF	2006	Positivos	É invariante à escala e rotação, e parcialmente insensitivo à variação de iluminação e oclusão parcial.
		Negativos	É computacionalmente mais pesado que os outros métodos, porém é mais rápido que o SIFT.

Fonte: Kelvin Salton do Prado, 2017.

Por meio do Quadro 2 pode-se visualizar que cada método possui características próprias e, por este motivo, é possível obter diferentes resultados nos experimentos utilizando o mesmo conjunto de dados.

3.3 Comitê de máquinas

O termo comitê de máquinas apresentado neste trabalho está associado à abordagem de *ensembles* que, basicamente, trata-se da geração de uma saída, neste caso, única,

baseando-se na saída de outras abordagens individuais, com o objetivo de melhorar a precisão de um problema de classificação ou predição.

Lima (2004) coloca que:

Um ensemble consiste de um conjunto de componentes, cada um representando uma proposta de solução, cujas classificações, no caso de problemas de classificação de padrões, ou estimativas, no caso de problemas de regressão, são combinadas de formas diversas visando um ganho de desempenho (LIMA, 2004).

Tratando-se de problemas de classificação, a abordagem mais comum para a combinação dos resultados dos componentes do *ensemble* é o voto majoritário (HANSEN; SALAMON, 1990), onde o resultado escolhido é aquele que obtém o maior número de votos. Para melhor compreender o funcionamento da abordagem, a Figura 16 apresenta um exemplo gráfico do esquema de funcionamento de um comitê de máquinas.

Figura 16 – Esquema de funcionamento do comitê de máquinas.

Fonte: Kelvin Salton do Prado, 2016.

Como pode ser visto na Figura 16, é possível utilizar quantos classificadores, também chamados de componentes do *ensemble*, forem necessários. Para que um *ensemble* de classificadores tenha um desempenho melhor do que os componentes independentes, é esperado que os componentes que o integram tenham um bom desempenho isoladamente e sejam diferentes entre si (HANSEN; SALAMON, 1990 apud LIMA, 2004). Neste caso, dois classificadores são considerados diferentes se eles apresentam erros de classificação diferentes em relação a um mesmo conjunto de dados (LIMA, 2004).

Algumas pesquisas relacionadas ao reconhecimento facial já mostraram que o uso de *ensembles* tende a melhorar a precisão dos classificadores (e.g. (GUTTA; WECHSLER, 1996; HUANG et al., 2000)). Por este motivo, nesse trabalho foi também implementada uma abordagem de comitê de máquinas, utilizando como componentes os métodos *eigenfaces*, *fisherfaces*, LBPH, SIFT e SURF, visando a melhorar o desempenho do processo de reconhecimento facial. Foram implementadas duas abordagens de comitês de máquinas, utilizando diferentes sistemas de votação, as quais são apresentadas a seguir.

3.3.1 Voto majoritário

A primeira abordagem de comitê de máquinas desenvolvida utiliza um sistema de voto majoritário. Nesta abordagem são obtidos os resultados de cada método de reconhecimento facial individualmente, e então é utilizado o voto majoritário para definir a qual sujeito a imagem facial pertence.

Considerando que todos os algoritmos foram treinados utilizando a mesma base de treinamento, dada uma nova imagem (e.g. Imagem Facial 1), cada método realiza o reconhecimento facial e apresenta um resultado (e.g. S1 - Sujeito 1), o qual será utilizado como entrada para o algoritmo de voto majoritário, que por sua vez irá combinar os resultados dos métodos para obter um novo resultado.

O Quadro 3 representa um exemplo prático do uso de comitê de máquinas utilizando o voto majoritário. Nele, é possível ver que os resultados apresentados pelo voto majoritário possuem uma chance maior de estarem corretos, pois são definidos com base nos resultados dos 5 métodos de reconhecimento facial.

Quadro 3 – Exemplo de comitê de máquinas utilizando voto majoritário

Imagens	Eigenfaces	Fisherfaces	LBPH	SIFT	SURF	Resultado
Imagem 1	S1	S5	S2	S1	S1	S1
Imagem 2	S4	S4	S3	S3	S3	S3
Imagem 3	S1	S4	S5	S2	S2	S2
Imagem 4	S5	S1	S5	S3	S3	S5

Fonte: Kelvin Salton do Prado, 2017.

O algoritmo de votação, por sua vez, recebe como parâmetro um vetor de inteiros com 5 posições, contendo os resultados de cada método de reconhecimento facial para uma determinada imagem. Em caso de empate no sistema de votação, o algoritmo irá

selecionar o resultado com base na ordem dos valores do vetor, assim como pode ser visto na linha correspondente a “Imagem 4” do Quadro 3.

3.3.2 Voto ponderado

A segunda abordagem de comitê de máquinas desenvolvida utiliza um sistema de voto ponderado. Nesta abordagem, são atribuídos pesos para cada componente (neste caso para cada método de reconhecimento facial) e então é realizada uma somatória de acordo com os resultados individuais de cada método, para definir a qual sujeito a imagem facial pertence.

O Quadro 4 apresenta um exemplo de utilização do sistema de voto ponderado, facilitando assim seu entendimento. Como é possível visualizar nesse exemplo, foram atribuídos pesos maiores para os métodos Fisherfaces e Eigenfaces, o que pode alterar o resultado final do comitê de máquinas, dependendo dos resultados dos outros métodos. Um exemplo disto pode ser visto na linha da Imagem 2, em que, embora o sistema de voto majoritário (Quadro 3) tenha classificado a imagem como sendo do sujeito 3 (S3), o sistema de voto ponderado (Quadro 4), com os parâmetros definidos, classificou a mesma imagem como sendo do sujeito 4 (S4). Neste caso, o resultado foi alterado devido aos pesos definidos para os algoritmos Fisherfaces e Eigenfaces. Cabe esclarecer que, neste exemplo, os pesos atribuídos e os resultados classificados pelos algoritmos foram definidos com o objetivo de exemplificar os possíveis resultados obtidos pela utilização do sistema de voto ponderado, demonstrando ainda casos em que houve empate, ou seja, o somatório dos pesos correspondentes a dois ou mais sujeitos foi equivalente e, como o método toma a decisão de selecionar apenas um dos sujeitos como resultado final da classificação, ele utiliza, neste caso, o sistema de voto majoritário explicado na seção anterior (Seção 3.3.1).

Quadro 4 – Exemplo de comitê de máquinas utilizando voto ponderado

Métodos	Eigenfaces	Fisherfaces	LBPH	SIFT	SURF	Somatório	Resultado
Pesos	25	30	10	15	20		
Imagem 1	S1	S5	S2	S1	S1	S1: 60 S2: 10 S5: 30	S1
Imagem 2	S4	S4	S3	S3	S3	S3: 45 S4: 55	S4
Imagem 3	S1	S4	S5	S2	S2	S1: 25 S2: 35 S4: 30 S5: 10	S2
Imagem 4	S2	S3	S5	S4	S5	S2: 25 S3: 30 S4: 15 S5: 30	S5
Imagem 5	S1	S2	S1	S3	S3	S1: 35 S2: 30 S3: 35	S1

Fonte: Kelvin Salton do Prado, 2017.

Desta forma é possível definir pesos mais altos para métodos que tenham um melhor desempenho individualmente, buscando melhorar os resultados obtidos. Acredita-se que se um método teve um desempenho individual maior que outros, ele é mais eficiente que os demais no ambiente em que os testes foram realizados, e assim poderá ter um peso maior na votação influenciando o resultado final.

Apesar de ser possível definir qualquer peso para cada um dos métodos, neste exemplo optou-se por manter um padrão onde a somatória de todos os pesos resultem em 100, pois facilita a visualização da percentagem de representação de cada método. Em caso de empate na votação do sistema ponderado, exemplo da linha correspondente a Imagem 4 do Quadro 4, o algoritmo irá utilizar o mesmo conceito do voto majoritário, ou seja, irá selecionar o resultado mais votado e, ainda em caso de empate, irá selecionar o resultado com base na ordem dos valores do vetor passado como parâmetro, exemplo da linha correspondente a Imagem 5 do Quadro 4.

Cabe ainda esclarecer que, caso seja utilizado o mesmo peso para todos os componentes do comitê, o sistema de voto ponderado irá funcionar da mesma maneira que o sistema de voto majoritário.

3.4 OpenCV

A OpenCV (Open Source Computer Vision Library)¹⁰ é uma biblioteca de visão computacional de código aberto e multiplataforma, originalmente desenvolvida pela Intel por volta do ano 2000, e livre tanto para uso acadêmico como para uso comercial. Essa biblioteca disponibiliza diversos algoritmos de visão computacional como, por exemplo, filtros de imagem, reconhecimento de objetos e reconhecimento facial, e possui ainda suporte a diversas linguagens, como Python, C++, Ruby, Matlab, entre outras (OPENCV, 2000; BRADSKI; KAEHLER, 2008).

O objetivo principal da biblioteca OpenCV é prover uma infraestrutura de visão computacional simples para auxiliar os desenvolvedores a construir aplicações sofisticadas de forma relativamente rápida (BRADSKI; KAEHLER, 2008). Por ser uma biblioteca de código aberto, multiplataforma e de fácil utilização, a OpenCV é bem difundida na área acadêmica, principalmente nas áreas de processamento de imagens e reconhecimento de padrões, sendo possível encontrar diversos trabalhos que fazem uso da mesma.

3.4.1 OpenCV contrib

O OpenCV Contrib é um repositório¹¹ criado para o desenvolvimento, armazenamento e compartilhamento de alguns módulos extras da OpenCV. Novos módulos que ainda não possuem uma interface de programação de aplicação (API) estável são disponibilizados neste repositório para, futuramente, quando o módulo evoluir e ganhar popularidade, serem movidos para o repositório central da OpenCV, e então serem disponibilizados oficialmente com a biblioteca.

Cabe aqui ressaltar que os algoritmos de detecção e reconhecimento facial utilizados neste trabalho foram fornecidos tanto pela biblioteca oficial da OpenCV quanto pelos módulos extras disponibilizados pelo OpenCV contrib.

¹⁰ Site oficial da biblioteca OpenCV: www.opencv.org

¹¹ Repositório OpenCV Contrib: [www.github.com/Itseez/opencv_contrib](https://github.com/Itseez/opencv_contrib)

3.5 Transformada Wavelet

Com base em Burrus, Gopinath e Guo (1998), Costa (2016) esclarece que:

A transformada Wavelet pode ser descrita como uma técnica de processamento de sinal, que realiza uma análise de um determinado sinal nos domínios do tempo e frequência, ao contrário da transformada de Fourier que realiza apenas a análise no domínio do tempo (BURRUS; GOPINATH; GUO, 1998 apud COSTA, 2016).

Ainda de acordo com Burrus, Gopinath e Guo (1997), uma onda é geralmente definida como uma função oscilante de tempo ou espaço, assim como uma senoide. Uma Wavelet pode ser considerada uma “onda pequena”, que tem sua energia concentrada no tempo. A Wavelet possui também a característica de onda oscilante, mas permite uma análise simultânea de tempo e frequência com uma fundação matemática flexível. A Transformada Discreta de Wavelet¹², por sua vez, descreve o sinal original e pode ser usada de forma semelhante aos coeficientes da série de Fourier para análises, descrição, aproximação e filtragem (BURRUS; GOPINATH; GUO, 1998).

Tendo como base as funções Wavelet, Mallat (2008) propôs uma nova forma de realizar a transformada Wavelet (COSTA, 2016). Nesta abordagem, são utilizados um conjunto de filtros, sendo um filtro passa-baixa g , o qual representa uma aproximação do sinal original, e outro filtro passa-alta h , o qual representa os detalhes da imagem original (COSTA, 2016).

Para a utilização da transformada Wavelet em sinais 2D, tais como imagens, a ideia original foi expandida para tratar as linhas e colunas como sinais independentes. Ao final da decomposição 2D são geradas 4 subimagens, onde cada uma delas apresenta dimensão $\frac{n}{2} \times \frac{n}{2}$, sendo que as imagens HL, LH e HH contém pelo menos as altas frequências em alguma das direções e podem ser descritas como contendo os detalhes horizontais, verticais e diagonais da imagem original, enquanto que a imagem LL representa uma aproximação desta (COSTA, 2016).

A Figura 17 representa os passos de decomposição da Wavelet 2D, onde são geradas as 4 subimagens LL (aproximação), LH (detalhes horizontais), HL (detalhes verticais), e HH (detalhes diagonais). Dada uma imagem de entrada, representada por I , primeiramente é aplicado um par de filtros passa-baixa e passa-alta em cada linha da imagem, o que resulta na diminuição da largura da mesma pela metade, para posteriormente, utilizando os dois resultados da imagem filtrada horizontalmente, ser aplicado outro par de filtros em

¹² Do inglês *Discrete Wavelet Transform* (DWT).

cada coluna da imagem, o que acaba por diminuir a altura da mesma. Este processo de diminuição das dimensões da imagem, chamado por alguns autores de *down-sampling*, é aplicado após cada filtro e é representado pelas setas para baixo na Figura 17 (POULARIKAS, 2010). Neste contexto, supondo que a imagem de entrada tivesse dimensões de 200×200 *pixels*, ao final seriam geradas 4 novas imagens de dimensões 100×100 *pixels*, originadas da aplicação dos filtros passa-baixa e passa-alta nas linhas e colunas da imagem original.

Figura 17 – Decomposição Wavelet 2D.

Fonte: Kelvin Salton do Prado, 2017.

Ao final, a transformada Wavelet gera um conjunto de coeficientes que podem ser utilizados para representar a imagem original, sendo que uma das formas mais comuns é utilizar os componentes da imagem LL, os quais representam uma aproximação da imagem original, apresentando menor dimensão e boa representatividade (BURRUS; GOPINATH; GUO, 1998 apud COSTA, 2016). A Figura 18 representa um exemplo das subimagens geradas pela transformada Wavelet 2D (LL, LH, HL e HH na Figura), bem como o resultado da operação inversa da transformada utilizando como parâmetro apenas a imagem LL (DWT2 inversa, na Figura), recuperando assim uma representação aproximada da imagem original.

Figura 18 – Resultado da Decomposição Wavelet 2D.

Fonte: Kelvin Salton do Prado, 2017.

Neste trabalho foi utilizada a transformada discreta de Wavelet 2D (função *dwt2*) disponibilizada pelo Matlab, para diminuir a dimensão das imagens, buscando de certa forma piorar a qualidade dessas com a intenção de aproximá-las à qualidade das imagens obtidas nos vídeos. Mais informações sobre a função *dwt2* podem ser conferidas na Seção 4.2.3.

3.6 Curva ROC

A curva ROC¹³, também chamada no português de Característica de Operação do Receptor (COR), foi desenvolvida no âmbito da detecção de sinais eletrônicos e problemas com radares durante a Segunda Guerra Mundial (ZWEIG; CAMPBELL, 1993 apud MARTINEZ; LOUZADA-NETO; PEREIRA, 2003), com o principal objetivo de quantificar a habilidade dos operadores dos radares em distinguir um sinal (e.g. um avião inimigo) de um ruído (e.g. uma ave) (REISER; FARAGGI, 1997 apud MARTINEZ; LOUZADA-NETO; PEREIRA, 2003). Nos anos 70, porém, a metodologia se disseminou em vários ramos da pesquisa biomédica, onde era utilizada, basicamente, no auxílio da classificação de indivíduos doentes e não doentes (MARTINEZ; LOUZADA-NETO; PEREIRA, 2003).

¹³ Do inglês *Receiver Operating Characteristic* (ROC).

Com isto, a curva ROC, já conhecida na literatura, representa um método para avaliar o desempenho de experimentos, incluindo a avaliação de problemas relacionados ao reconhecimento de padrões e ao aprendizado de máquina. Diferente das medidas de precisão e revocação (apresentadas na Seção 3.7), a curva ROC tem como objetivo examinar a relação entre a taxa de verdadeiros positivos (do inglês *True Positive Rate* - TPR) e a taxa de falsos positivos (do inglês *False Positive Rate* - FPR) (CHAO, 2007).

Neste caso, para traçar a curva ROC é preciso primeiramente calcular a matriz de confusão, também chamada de matriz de classificação, para cada limiar ou parâmetro do sistema. Neste contexto, o limiar corresponde à medida de distância e é responsável por representar o resultado da classificação, ou seja, para os resultados abaixo do limiar pode-se considerar que o algoritmo reconheceu o sujeito e para os resultados acima do limiar que o algoritmo não reconheceu. Cabe esclarecer ainda que o valor do limiar é definido pelo usuário, e quanto mais baixo for o limiar maior será a precisão do algoritmo, pois apenas imagens com distâncias muito próximas serão consideradas como reconhecidas. Porém, por outro lado, utilizando um limiar baixo pode-se, de certa forma, perder informações que poderiam ser relevantes. Antes de apresentar o esquema da matriz de confusão utilizada pela curva ROC, cabe apresentar os possíveis resultados que podem ser obtidos por meio da utilização do limiar:

- Resultados abaixo do limiar:
 - Reconhecido (Verdadeiro Positivo): o sujeito correspondente à imagem foi corretamente classificado e a confiança está abaixo do limiar.
 - Não reconhecido (Falso Positivo): o sujeito correspondente à imagem foi incorretamente classificado porém a confiança está abaixo do limiar.
 - Não é imagem facial ou não é conhecida (Falso Positivo): não é uma imagem facial ou não corresponde a um sujeito encontrado na base de treinamento e a confiança está abaixo do limiar.
- Resultados acima do limiar:
 - Reconhecido (Falso Negativo): o sujeito correspondente à imagem foi corretamente classificado porém a confiança está acima do limiar.
 - Não reconhecido (Verdadeiro Negativo): o sujeito correspondente à imagem foi incorretamente classificado e a confiança está acima do limiar.

- Não é imagem facial ou não é conhecida (Verdadeiro Negativo): não é uma imagem facial ou não corresponde a um sujeito encontrado na base de treinamento e a confiança está acima do limiar.

Para facilitar o entendimento, a Figura 19 apresenta de forma gráfica cada um dos possíveis resultados que podem ser obtidos quando utilizando o limiar.

Figura 19 – Possíveis resultados utilizando o limiar.

Fonte: Kelvin Salton do Prado, 2017.

Como é possível notar na Figura 19, cada imagem correspondente a um resultado possui um símbolo verde ou vermelho no canto inferior direito, o qual representa respectivamente se o resultado está correto ou incorreto, sendo que o ideal seria que abaixo do limiar estivessem apenas imagens faciais corretamente reconhecidas e que acima do limiar estivessem as imagens incorretamente reconhecidas ou imagens apontadas como sendo faces pelo algoritmo de detecção facial, mas que na verdade não correspondiam a uma face. Além disso, acima de cada imagem encontra-se o nome do grupo ao qual o resultado pertence (e.g. Verdadeiro Positivo (VP), Falso Negativo (FN)).

Tendo compreendido a responsabilidade do limiar, bem como os possíveis resultados que podem ser obtidos pelos algoritmos de reconhecimento facial, pode-se então apresentar

o esquema da matriz de confusão (Figura 20), a qual é utilizada para a geração da curva ROC e para o cálculo das medidas de precisão e revocação apresentadas na Seção 3.7.

Figura 20 – Esquema da matriz de confusão utilizada pela curva ROC.

		Condições da Predição	
		Predição Positiva (Abaixo do Limiar)	Predição Negativa (Acima do Limiar)
Condições Verdadeiras	População Total		
	Condição Positiva	VERDADEIRO POSITIVO (VP) Imagens faciais reconhecidas corretamente (abaixo do limiar).	FALSO NEGATIVO (FN) Imagens faciais reconhecidas corretamente (acima do limiar).
	Condição Negativa	FALSO POSITIVO (FP) Imagens reconhecidas incorretamente (abaixo do limiar).	VERDADEIRO NEGATIVO (VN) Imagens que não deveriam ser reconhecidas (acima do limiar).

Fonte: Kelvin Salton do Prado, 2017.

A partir da matriz de confusão é possível calcular a taxa de verdadeiros positivos e a taxa de falsos positivos, que são utilizadas na geração da curva ROC, por meio das seguintes fórmulas (FAWCETT, 2006):

$$\text{Taxa_VP} = \frac{VP}{VP + FN}, \quad (28)$$

$$\text{Taxa_FP} = \frac{FP}{FP + VN} \quad (29)$$

Por meio dos estudos iniciais realizados, incluindo a revisão sistemática, foi possível encontrar diversos trabalhos na área de reconhecimento facial que fazem uso da curva ROC para medir o desempenho dos experimentos (e.g. (PENTLAND et al., 1994; SUGANYA; MENAKA, 2014; RIBEIRO; CHIACHIA; MARANA, 2010)). Por este motivo, utilizou-se este método, juntamente com as medidas de precisão e revocação, apresentadas a seguir, para medir o desempenho dos experimentos deste trabalho.

3.7 Precisão e revocação

Precisão e revocação são medidas conhecidas na área de reconhecimento de padrões para mensurar a relevância dos resultados obtidos por algum método ou algoritmo específico. De acordo com Buckland e Gey (1994), a medida de precisão¹⁴ corresponde ao número de itens relevantes selecionados, em relação ao número total de itens selecionados, sendo portanto uma medida de pureza para o método de seleção, ou até uma medida de eficácia na exclusão de itens não relevantes do conjunto selecionado. O valor ideal ou desejável para a precisão é sempre de 100%, pois neste caso o método selecionou ou recuperou apenas itens relevantes.

Resumidamente, a precisão mede basicamente o quão útil ou correto os resultados obtidos estão, comparando o total de itens relevantes selecionados em relação ao total de itens selecionados ou recuperados. Tendo como base os parâmetros apresentados na Figura 20 da Seção 3.6 é possível calcular a precisão utilizando a seguinte equação (FAWCETT, 2006):

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (30)$$

Buckland e Gey (1994) também esclarecem que a medida de revocação¹⁵ corresponde ao número de itens relevantes selecionados em relação ao total de itens relevantes, e pode ser vista como uma medida de eficácia na seleção de itens relevantes de um conjunto. Nem sempre uma alta taxa de revocação é necessária, pois em alguns casos a pessoa pode querer selecionar apenas alguns itens relevantes (BUCKLAND; GEY, 1994). Por exemplo, suponha que uma pessoa esteja fazendo uma pesquisa em uma grande base de dados, e que esta pessoa deseja selecionar itens relacionados à palavra “fruta”. Neste caso, é possível que tal pessoa não queira selecionar todos os itens relacionados a fruta, e sim apenas alguns, de preferência os mais relevantes. Sendo assim, não é necessário obter uma revocação de 100%. Através disto, é possível concluir que a medida de revocação mede basicamente o quão completos os resultados estão em relação ao total de itens, incluindo os itens relevantes não selecionados. Neste contexto, utilizando como base os parâmetros apresentados na Figura

¹⁴ Do inglês, *precision*.

¹⁵ Do inglês, *recall*.

20 da Seção 3.6, é possível calcular a revocação por meio da seguinte equação (FAWCETT, 2006):

$$\text{Revocação} = \frac{VP}{VP + FN} \quad (31)$$

Note que a Equação 31 corresponde à taxa de verdadeiros positivos e, por este motivo, é igual a Equação 28.

Para facilitar o entendimento, consideremos o seguinte exemplo: supondo que em um determinado vídeo foi possível detectar 10 imagens de faces, das quais o algoritmo de reconhecimento facial conseguiu reconhecer 8 (abaixo do limiar - $(VP + FP)$). Destas 8 imagens o algoritmo reconheceu corretamente apenas 5 delas (VP) e as 3 restantes foram reconhecidas incorretamente (FP). As 2 imagens restantes foram reconhecidas corretamente, porém acima do limiar¹⁶. Sendo assim, tem-se os seguintes parâmetros:

- Verdadeiro Positivo (VP): 5
- Falso Positivo (FP): 3
- Falso Negativo (FN): 2
- Verdadeiro Negativo (VN): 0

Desta forma é possível calcular e obter os seguintes resultados de precisão e revocação:

$$\text{Precisão} = \frac{5}{5 + 3} = 0,6250 \quad (32)$$

$$\text{Revocação} = \frac{5}{5 + 2} = 0,7143 \quad (33)$$

Pode-se ainda multiplicar os resultados por 100 para obter os resultados em percentagem. Sendo assim foi possível obter uma precisão de 62.50% e uma revocação de 71.43%.

As medidas de precisão e revocação, assim como a curva ROC, serão utilizadas para medir os resultados obtidos nos experimentos, servindo como uma ferramenta auxiliar para alcançar os objetivos deste trabalho.

¹⁶ Tratam-se de falsos negativos, ou seja, o algoritmo realizou a classificação corretamente, porém o valor de confiança está acima do limiar, como demonstrado na Figura 19 apresentada anteriormente.

3.8 Interpolação Bicúbica

O princípio básico do processo de redimensionar uma imagem, aumentando as suas dimensões, também chamado no inglês de *image magnification* ou *upsampling*, consiste basicamente em aumentar o número de *pixels*, incluindo novos *pixels* com base nos *pixels* existentes na imagem original. Métodos de interpolação são frequentemente utilizados no redimensionamento de imagens. Dentre os diversos métodos pode-se citar: a interpolação do vizinho mais próximo, a interpolação bilinear e a interpolação bicúbica (HAN, 2013).

Antes de apresentar o método de interpolação bicúbica é preciso primeiramente apresentar o método de interpolação bilinear. Enquanto a interpolação do vizinho mais próximo basicamente replica o valor do *pixel* mais próximo (Figura 21), sem considerar os valores dos demais *pixels* vizinhos, a interpolação bilinear (extensão da interpolação linear) utiliza uma média ponderada para determinar o valor de um novo *pixel*, com base nos 4 *pixels* vizinhos (2×2) da imagem original (Figura 22 (b)), utilizando para isso, o produto de duas funções lineares, aplicadas nas linhas e colunas da imagem (Figura 22 (c) e (d)). O método de interpolação bicúbica, por sua vez, utiliza o mesmo princípio da interpolação bilinear, porém é mais sofisticado e produz bordas mais suavizadas, se comparado ao método de interpolação bilinear. Na interpolação bicúbica, um novo *pixel* utiliza os 16 *pixels* vizinhos (4×4) ao *pixel* da imagem original, como pode ser visualizado na Figura 23 (b) (PRASHANTH; SHASHIDHARA; KN, 2009; DOURADO, 2014). A interpolação bicúbica é uma extensão da interpolação cúbica destinada a realizar a interpolação em dados bidimensionais, como por exemplo em imagens. Cabe esclarecer que, nas Figuras 21, 22 e 23, os valores da intensidade dos *pixels* são aproximados, e devem ser considerados apenas para o entendimento de cada método de interpolação.

Figura 21 – Exemplo de interpolação utilizando o método do vizinho mais próximo.

Fonte: Kelvin Salton do Prado, 2017.

Figura 22 – Exemplo de interpolação utilizando o método bilinear.

Fonte: Kelvin Salton do Prado, 2017.

Figura 23 – Exemplo de interpolação utilizando o método bicúbico.

Fonte: Kelvin Salton do Prado, 2017.

Com a intenção de facilitar o entendimento, consideremos o seguinte exemplo gráfico dos métodos de interpolação em um conjunto de dados unidimensional. Supondo que se tem uma imagem em escala de cinza de dimensões 1×10 *pixels*, a qual pode ser visualizada na Figura 24, a intensidade dos *pixels* pode ser representada em forma de um gráfico, assim como é demonstrado no gráfico superior esquerdo da Figura 25.

Figura 24 – Exemplo de imagem 1×10 *pixels*.

Fonte: Kelvin Salton do Prado, 2017.

Com base nos dados da imagem original é possível aplicar os métodos de interpolação de vizinho mais próximo, linear e cúbica, com o objetivo de redimensionar os dados da imagem para uma dimensão de 1×37 *pixels*, ou seja, incluindo novos *pixels* com base nos

pixels da imagem original. Neste caso, os dados foram redimensionados de 10 pontos para 37 pontos e os resultados são apresentados nos gráficos da Figura 25.

Figura 25 – Exemplo dos principais métodos de interpolação.

Fonte: Kelvin Salton do Prado, 2017.

Como é possível visualizar nos gráficos apresentados na Figura 25, o método de interpolação cúbica gera uma curva mais suavizada, se comparado aos outros dois métodos de interpolação apresentados. Com isso, é possível concluir que o resultado de uma imagem redimensionada utilizando a interpolação bicúbica (extensão da interpolação cúbica) será mais suavizada, e consequentemente terá um resultado visivelmente melhor e mais próximo da imagem original.

Ainda, cabe colocar que o método de interpolação utilizado no redimensionamento de imagens digitais pode ser escolhido de acordo com o objetivo geral da aplicação, ou seja, dependendo do objetivo geral que deseja-se alcançar, o método de interpolação do vizinho mais próximo pode até ser uma boa escolha, porém de maneira geral pode-se dizer que o método de interpolação bicúbica produz bons resultados, se comparado a outros métodos de interpolação (e.g. interpolação do vizinho mais próximo, Figura 26 (b)), pois suaviza as bordas da imagem, gerando uma imagem relativamente de boa qualidade e visualmente semelhante à imagem original, como pode ser visualizado no exemplo da Figura 26 (d). Além disso, a interpolação bicúbica é o método mais comumente utilizado em softwares de edição de imagens para o redimensionamento de imagens (PRASHANTH; SHASHIDHARA; KN, 2009).

Figura 26 – Exemplo dos principais métodos de interpolação.

Fonte: Kelvin Salton do Prado, 2017.

Com base nos motivos apresentados nesta seção, a interpolação bicúbica foi utilizada na etapa de pré-processamento dos dados, onde as imagens são redimensionadas para um tamanho padrão, que é utilizado tanto na etapa de treinamento como na etapa de teste.

4 Experimentos

Neste capítulo serão apresentados os experimentos realizados, buscando facilitar o entendimento do trabalho como um todo e esclarecer futuramente os resultados e conclusões obtidos.

Para compreender o processo de reconhecimento facial realizado, e consequentemente entender como os experimentos foram conduzidos, consideremos a Figura 27, que representa de forma gráfica os processos realizados. Nela, os três primeiros quadros (parte superior da Figura 27) estão relacionados à etapa de detecção facial, na qual é aplicado o algoritmo Viola-Jones em cada *frame* do vídeo, com o objetivo de extrair todas as imagens faciais encontradas.

Figura 27 – Esquema do processo de reconhecimento facial.

Fonte: Kelvin Salton do Prado, 2017.

A partir das imagens extraídas dos vídeos, é realizada uma classificação manual, alterando o nome do arquivo da imagem de acordo com o número correspondente a cada

participante, definidos na base de treinamento. Os três últimos quadros (parte inferior da Figura 27) correspondem à etapa de reconhecimento facial, na qual os algoritmos são treinados utilizando as bases de treinamento, descritas na Seção 4.2.3, e é realizada a classificação de cada face encontrada nas bases de teste, conforme descrito na Seção 4.2.4.

4.1 Informações técnicas

Os experimentos foram realizados em uma máquina com um processador Intel Core i5 e 6 gigabytes de memória de acesso aleatório¹, utilizando o sistema operacional Ubuntu 14.04 e as versões 2.7.6 do Python e 3.1.0 da biblioteca OpenCV. As câmeras de segurança utilizadas na captura dos vídeos nos experimentos são mini câmeras *color* modelo N-150 da marca Nitrix. Um exemplo deste modelo pode ser visualizada na Figura 28:

Figura 28 – Modelo de câmera utilizada nos experimentos.

Fonte: Kelvin Salton do Prado, 2017.

Apesar da qualidade de imagem da câmera ser relativamente boa, se comparada a outras câmeras de mesma categoria, ela não é uma câmera ideal para capturar imagens com o propósito de realizar reconhecimento facial, pois trata-se de uma câmera de vigilância, as quais são geralmente instaladas em um ponto fixo para obter uma imagem mais ampla de um ambiente, ou seja, sem ter a intenção de capturar exatamente a face frontal do sujeito.

¹ Do inglês Random Access Memory (RAM).

4.2 Dados coletados

4.2.1 Academia Central Kungfu-Wushu

Nesta seção será apresentada uma breve descrição sobre a academia Central Kungfu-Wushu, a qual disponibilizou todos os dados necessários (imagens e vídeos) para a realização deste trabalho.

A academia Central Kungfu-Wushu² está localizada na cidade de Campinas, estado de São Paulo, e oferece aulas de diversas artes marciais, dentre elas, o Kung-fu tradicional, a Sanda (conhecido também como Boxe-Chinês) e o Tai Chi Chuan.

Além disto, a academia, coordenada pelo professor Enrique Ortega, disponibiliza aulas para crianças, jovens e adultos e oferece também treinamentos específicos para um grupo que participa de campeonatos nacionais e internacionais.

Os 18 participantes deste trabalho fazem parte deste grupo específico, e foram selecionados com o objetivo de limitar e ter um maior controle sobre os experimentos realizados.

4.2.2 Participantes da pesquisa

Como mencionado, os experimentos foram realizados com um grupo específico da academia Central Kungfu-Wushu, e por este motivo apenas os dados destes participantes foram coletados para serem utilizados no treinamento dos algoritmos. O grupo de participantes voluntários desta pesquisa é conhecido como equipe de competição pois, além dos treinos normais, eles realizam um treinamento diferenciado para a participação em competições. O grupo é composto por 18 pessoas, sendo 4 do sexo feminino e 14 do sexo masculino, com idades variadas.

Como a academia Central Kungfu-Wushu possui diversos alunos, foi necessário reduzir o número de participantes para a realização dos experimentos, para que desta forma se obtenha um controle maior sobre os experimentos realizados. Acredita-se que com um grupo de 18 participantes bem diversificados, incluindo adolescentes e adultos do sexo feminino e masculino, é possível obter resultados suficientes para alcançar os objetivos da pesquisa.

² Site da academia Central Kungfu-Wushu: www.centralkungfu.com.br

Como o objetivo da pesquisa era ser o menos intrusivo possível, optou-se por realizar os experimentos coletando apenas as imagens da ficha de cadastro dos alunos, as quais foram utilizadas na etapa de treinamento dos algoritmos de reconhecimento facial. Com isto criou-se a primeira base de treinamento, contendo uma imagem de pose frontal de cada participante, e, a partir disso, criaram-se as demais bases de treinamento que são detalhadas na Seção 4.2.3.

4.2.3 Bases de treinamento

Para realizar o processo de reconhecimento facial, primeiramente é preciso treinar os algoritmos com as imagens faciais dos sujeitos que pretende-se identificar. Nesta seção serão apresentadas as bases de imagens utilizadas para o treinamento dos algoritmos.

Os experimentos foram realizados com um total de 18 participante de idades e gêneros diferentes. A primeira etapa foi composta pela coleta das imagens da ficha de cadastro da academia, as quais possuem uma imagem do tronco e da cabeça de cada participante, como pode ser visto no exemplo da Figura 29.

Figura 29 – Imagens originais da ficha de cadastro da academia.

Fonte: Kelvin Salton do Prado, 2017.

A partir dessas imagens, foi realizada a extração das imagens faciais utilizando o algoritmo Viola-Jones descrito na Seção 3.1.1. Com isto, foi possível extrair as imagens faciais de cada um dos participante, como pode ser visto na Figura 30, as quais foram utilizadas como base para a criação de todas as bases de treinamento que serão descritas a seguir.

Figura 30 – Imagens faciais extraídas da ficha de cadastro da academia.

Fonte: Kelvin Salton do Prado, 2017.

A maior parte das bases de treinamento criadas neste trabalho utilizam a transformada Wavelet discreta para gerar imagens com uma qualidade mais baixa, com a intenção de tentar aproximar essas imagens à qualidade das imagens obtidas por meio dos vídeos de teste. Para aplicar a transformada Wavelet discreta foi utilizada a função *dwt2* do software Matlab³, a qual aplica uma decomposição Wavelet bidimensional de nível único. A função *dwt2* disponibilizada pelo Matlab recebe um parâmetro, chamado de *wname*, o qual representa o nome da Wavelet utilizada na decomposição.

Na criação das bases de treinamento foram utilizados os parâmetros *haar* e *db2* da função *dwt2* do Matlab. Ambas as *wavelets*, *haar* (também conhecida como *db1*) e *db2*, pertencem à família das *wavelets* Daubechies. Essas *wavelets* foram utilizadas por serem conhecidas e difundidas na literatura (e.g. (MULCAHY, 1997; COSTA, 2016)). Porém existem diversas outras que poderiam ser utilizadas, podendo-se obter resultados diferentes.

Para essa pesquisa, foram criadas as seguintes bases de treinamento:

- Base 1: Criada apenas a partir das imagens do cadastro dos alunos, ou seja, foi utilizada uma imagem para cada participante, totalizando 18 imagens. Um exemplo desta base de treinamento pode ser visualizado na Figura 30.
- Base 2: Além de contar com as imagens do cadastro dos alunos, essa base também utiliza a transformada Wavelet discreta para gerar novas imagens com uma qualidade mais baixa. Nesta base de treinamento foi aplicado o primeiro nível da Wavelet *db2*,

³ Matlab: <https://www.mathworks.com/products/matlab.html>

- gerando ao final 2 imagens por participante, sendo uma da ficha de cadastro e outra do resultado da aplicação da Wavelet *db2*, o que resultou em um total de 36 imagens.
- Base 3: Criada utilizando as mesmas imagens da segunda base de treinamento, essa base inclui também um novo conjunto de imagens, geradas através da aplicação da transformada Wavelet discreta de segundo nível, utilizando como parâmetro a Wavelet *db2*, assim como na segunda base de treinamento. Nessa base, são utilizadas 3 imagens para cada participante, totalizando 54 imagens.
 - Base 4: Assim como na segunda base de treinamento, nessa base, além de serem utilizadas as imagens obtidas do cadastro dos alunos, também é aplicada a transformada Wavelet discreta de primeiro nível nas imagens do cadastro. Porém neste caso utiliza-se como parâmetro (*wname*) a Wavelet *haar*, gerando assim 2 imagens faciais de cada participante, sendo uma imagem da ficha de cadastro e outra gerada pela aplicação da Wavelet *haar*. Nesta base foi utilizado um total de 36 imagens.
 - Base 5: Semelhante a terceira base de imagens, a quinta base de imagens utiliza as imagens geradas na quarta base, porém inclui também a transformada Wavelet discreta de segundo nível, utilizando como parâmetro a Wavelet *haar*, assim como na quarta base de treinamento. Esta base possui 3 imagens faciais de cada participante, totalizando 54 imagens.
 - Base 6: A sexta e última base de treinamento faz uso de todas as imagens geradas pelas outras bases de treinamento. Além das imagens originais do cadastro dos alunos, ela utiliza todas as imagens geradas através da transformada Wavelet discreta, aplicando tanto a Wavelet *db2* quanto a *haar* nos níveis 1 e 2. Dessa forma, esta base de treinamento conta com 5 imagens faciais de cada participante, totalizando 90 imagens. Um exemplo das imagens geradas para esta base de treinamento pode ser visualizado na Figura 31, onde a primeira coluna (a) representa a imagem original da ficha de cadastro, a segunda (b) e terceira (c) colunas representam, respectivamente, o primeiro e segundo nível da Wavelet *db2* e a quarta (d) e quinta (e) colunas representam o primeiro e segundo nível da Wavelet *haar*.

Figura 31 – Exemplo de imagens faciais utilizadas na sexta base de treinamento.

Fonte: Kelvin Salton do Prado, 2017.

Com a intenção de facilitar o entendimento em relação as bases de treinamento criadas, o Quadro 5 mostra um resumo do que foi apresentado acima, descrevendo o tipo e a quantidade das imagens utilizadas em cada base.

Quadro 5 – Resumo das bases de treinamento

Base	Descrição das Imagens Utilizadas	Quantidade de Imagens
Base 1	Imagens da ficha de cadastro	18
Base 2	Imagens da ficha de cadastro Primeiro nível da Wavelet <i>db2</i>	36
Base 3	Imagens da ficha de cadastro Primeiro nível da Wavelet <i>db2</i> Segundo nível da Wavelet <i>db2</i>	54
Base 4	Imagens da ficha de cadastro Primeiro nível da Wavelet <i>haar</i>	36
Base 5	Imagens da ficha de cadastro Primeiro nível da Wavelet <i>haar</i> Segundo nível da Wavelet <i>haar</i>	54
Base 6	Imagens da ficha de cadastro Primeiro nível da Wavelet <i>db2</i> Segundo nível da Wavelet <i>db2</i> Primeiro nível da Wavelet <i>haar</i> Segundo nível da Wavelet <i>haar</i>	90

Fonte: Kelvin Salton do Prado, 2017.

Cabe esclarecer que optou-se por aplicar a transformada Wavelet até o segundo nível, pois acredita-se que, a partir deste, a qualidade das imagens geradas não é relevante

para os experimentos, devido à dificuldade de distinguir visualmente a qual sujeito a imagem pertence.

Como é possível visualizar na Figura 30 apresentada anteriormente, a qualidade das imagens faciais obtidas da ficha de cadastro é relativamente boa se comparada com as imagens faciais obtidas dos vídeos (Figura 32). Acredita-se que estas seriam imagens ideais para realizar o reconhecimento facial, pois possuem uma boa qualidade, uma iluminação adequada e uma pose frontal. Na Figura 32 é possível notar alguns pontos desfavoráveis para o reconhecimento facial, tais como variação de pose e baixa resolução das imagens.

Figura 32 – Amostra das imagens faciais obtidas do próprio vídeo.

Fonte: Kelvin Salton do Prado, 2016.

Cabe esclarecer que, como as bases de treinamento foram criadas a partir das imagens faciais do cadastro dos alunos, ou seja, apenas 1 imagem facial de pose frontal de cada participante, acredita-se que isto pode ser um gargalo no trabalho, sendo que o ideal seria obter mais imagens de treinamento com diferentes poses de cada participante. Apesar disto, o objetivo do trabalho é ser o menos intrusivo possível em seu ambiente de teste, utilizando apenas os dados já existentes. Neste caso, pode-se sugerir como trabalho futuro criar uma ou mais bases de treinamento mais robustas, com um número maior de imagens de cada participante, preferencialmente com variações de pose, iluminação e nitidez.

Apesar de as *wavelets* serem utilizadas na criação das bases de treinamento, não pretende-se aprofundar os estudos sobre elas, pois acredita-se que este é um campo relativamente amplo e, por este motivo, foge ao escopo deste trabalho.

Por fim, foram realizados experimentos com as diversas bases de treinamento nos quais encontraram-se algumas possíveis limitações do presente trabalho. Os resultados dos experimentos realizados e as possíveis limitações serão apresentados detalhadamente nas seções seguintes.

4.2.4 Bases de teste

Para a realização dos experimentos foram coletados 12 vídeos de 3 salas distintas da academia Central Kungfu-Wushu. No Quadro 6 é possível visualizar um resumo dos dados coletados, dentre eles o tempo total dos vídeos, a quantidade total de *frames* analisados e a quantidade total de imagens detectadas, sendo que a mesma foi gerada após a realização da detecção facial em todos os vídeos.

Quadro 6 – Resumo dos dados coletados da academia Central Kungfu-Wushu

Dados coletados	
Quantidade de vídeos coletados	12
Tempo total dos vídeos	08:58:29
Quant. total de <i>frames</i> analisados	821.104
Tempo médio dos vídeos	00:44:52
Quant. média de <i>frames</i> por vídeo	68.425
Quant. total de imagens detectadas	13.125
Quant. total de imagens faciais detectadas	3.360
Quant. total de imagens não faciais detectadas	9.765
Imagem facial de menor dimensão	21×21 pixels
Imagem facial de maior dimensão	87×87 pixels
Média da dimensão das imagens faciais	42×42 pixels

Fonte: Kelvin Salton do Prado, 2017.

Ainda, apesar de o Quadro 6 apresentar um resumo geral de todos os 12 vídeos coletados após a realização da detecção facial, utilizando o algoritmo Viola-Jones, 3 destes vídeos (vídeos 4, 7 e 8) foram removidos dos testes, restando um total de 9 vídeos. Os 3 vídeos foram removidos por não ter sido possível detectar um número aceitável de imagens faciais para realizar as próximas etapas dos experimentos. Cabe esclarecer que, o termo “aceitável” empregado neste parágrafo está relacionado a uma quantidade mínima de imagens faciais que foi, neste trabalho, definida como 10%. Ou seja, vídeos em que a quantidade de imagens faciais detectadas estava abaixo de 10%, em relação ao total de imagens detectadas no próprio vídeo⁴, foram removidos dos próximos experimentos. Além disso, cabe colocar que nos vídeos 4 e 8 não foi possível detectar qualquer imagem facial, e no vídeo 7 foi possível detectar apenas 2 imagens faciais (os resultados podem ser vistos na Tabela 2 da Seção 4.3.3).

⁴ O total de imagens detectadas no vídeo corresponde a soma das imagens faciais conhecidas (pertencentes aos participantes da pesquisa), das imagens faciais não conhecidas (de pessoas não participantes da pesquisa) e de imagens não faciais.

4.2.5 Questões éticas

Como os experimentos realizados neste trabalho fazem uso de imagens faciais de diversas pessoas, foi necessária uma aprovação prévia por parte dos participantes, com a intenção de viabilizar futuras publicações. A autorização foi obtida por meio do termo de consentimento livre e esclarecido (Apêndice D), no qual o participante autoriza a utilização de sua imagem para uso na pesquisa científica, dentro deste trabalho e demais publicações relacionadas ao mesmo. No caso de participante criança ou adolescente foi necessário também a aprovação prévia de um responsável, por meio do termo de assentimento (Apêndice E).

Nas seções seguintes serão apresentados os experimentos realizados e os resultados obtidos tanto na etapa de detecção facial quanto na etapa de reconhecimento facial. Cabe enfatizar que a detecção e o reconhecimento facial são duas etapas distintas, portanto os experimentos serão apresentados separadamente nas próximas seções.

4.3 Detecção facial

Como descrito anteriormente, a detecção facial é a primeira etapa do processo de reconhecimento facial, a qual é responsável por detectar e extrair as imagens faciais dos vídeos utilizados nos testes. Nas seções seguintes serão apresentados os parâmetros utilizados na etapa de detecção facial, bem como os resultados obtidos nesta etapa.

4.3.1 Descrição dos parâmetros

Para a detecção facial, foi aplicada a função conhecida como *detectMultiScale* da biblioteca OpenCV que, por sua vez, foi desenvolvida com base no trabalho de Viola e Jones (2001). Essa é uma função geral utilizada para a detecção de objetos, inclusive faces (OPENCV, 2000). Os parâmetros da função utilizados no desenvolvimento do algoritmo aplicado nesse trabalho foram:

- ***image***: a imagem na qual deseja-se detectar as faces, deve ser em escala de cinza.

No caso deste trabalho, a imagem passada como parâmetro é cada *frame* do vídeo.

- ***scaleFactor***: parâmetro que especifica quanto o tamanho da imagem será reduzido em cada escala da imagem. Como o modelo definido durante o treinamento possui um tamanho fixo, e na imagem de entrada as faces podem ter tamanhos variados, dependendo da distância entre a face e a câmera, este parâmetro é utilizado para redimensionar a imagem de entrada, compensando esta variação e tornando as faces detectáveis pelo algoritmo.
- ***minNeighbors***: parâmetro que especifica quantos vizinhos cada retângulo candidato deve ter para retê-lo. Como o algoritmo utiliza uma “janela móvel” para detectar os objetos, este parâmetro define quantos objetos são detectados próximos do atual antes de declarar a face encontrada.
- ***minSize***: define o tamanho mínimo do objeto a ser considerado. Os objetos detectados que tiverem tamanho menor que este parâmetro serão ignoradas.

Estes são alguns dos parâmetros aceitos pela função *detectMultiScale* da OpenCV (OPENCV, 2000). Os valores utilizados como parâmetros nos testes serão detalhados na Seção 4.3.2.

4.3.2 Seleção dos parâmetros

Foram utilizados 4 arquivos no formato *Extensible Markup Language* (XML) disponibilizados pela biblioteca OpenCV⁵ como classificadores em cascata pré-treinados⁶ para a realização da detecção facial, sendo que cada classificador foi testado com diferentes parâmetros.

Os experimentos foram realizados com os seguintes arquivos de entrada do classificador em cascata: *Haar-cascade frontal face default*, *LBP cascade frontal face*, *Haar-cascade frontal face alternative 2* e *Haar-cascade frontal face alternative 3*. Tais arquivos encontram-se no formato *Extensible Markup Language* (XML) e possuem características pré-treinadas que são utilizadas para detectar as faces. Cabe ainda colocar que todos os arquivos XML utilizados são distribuídos juntamente com o projeto da OpenCV e que, além dos arquivos utilizados para a detecção facial, existem outros arquivos de classificadores pré-treinados

⁵ Link para o repositório: <https://github.com/opencv/opencv/tree/master/data>.

⁶ Como a OpenCV é um projeto *open source*, os arquivos pré-treinados foram gerados pelos mantenedores do projeto com auxílio da comunidade em geral.

para detectar, por exemplo, o corpo de uma pessoa, o sorriso e até mesmo objetos ou animais.

Dos classificadores testados, o que teve maior número de imagens detectadas foi o *Haar-cascade frontal face default* com os seguintes parâmetros:

Quadro 7 – Parâmetros do algoritmo de detecção facial

XML: Haar-cascade frontal face default		
scaleFactor	minNeighbors	minSize
1.05	5	30

Fonte: Kelvin Salton do Prado, 2017.

Tal experimento detectou ao todo 3.588 imagens no vídeo, sendo 2.285 falsas faces, as quais não são realmente imagens faciais, e 1.303 faces verdadeiras. Neste caso, foram contabilizadas até mesmo faces com a nitidez extremamente baixa. Apesar deste experimento detectar várias imagens faciais verdadeiras (cerca de 36.32% do total), ele também detecta muitas faces falsas (cerca de 63.68% do total), ou seja, detecta mais imagens faciais falsas do que verdadeiras, o que prejudica o desempenho do reconhecimento facial, principalmente diminuindo sua velocidade e aumentando seu custo computacional, pois é necessário processar diversas imagens irrelevantes.

Já o classificador que teve o menor número de detecções faciais foi o *LBP cascade frontal face* e os parâmetros:

Quadro 8 – Parâmetros do algoritmo de detecção facial

XML: LBP cascade frontal face		
scaleFactor	minNeighbors	minSize
1.2	5	10

Fonte: Kelvin Salton do Prado, 2017.

Este experimento detectou apenas 215 imagens no vídeo, sendo 21 faces falsas (9.77% do total) e 194 faces verdadeiras (90.23% do total). Analisando apenas as porcentagens, pode-se concluir que os resultados foram de certa forma satisfatórios, pois a taxa de faces falsas foi baixa, se comparada ao experimento anterior. Por outro lado, ao verificar a quantidade de imagens detectadas, bem como a quantidade de faces falsas e verdadeiras, é possível visualizar claramente a diferença em relação ao experimento anterior, no qual foi possível detectar um total de 3.588 imagens, e concluir que este experimento não obteve um resultado satisfatório, pois não foi possível detectar uma boa parte das imagens faciais do vídeo.

A seguir serão apresentados dois experimentos que, na opinião do autor, tiveram os melhores resultados em comparação com os demais, pois detectaram uma quantidade de imagens considerada aceitável, ou seja, foi possível detectar menos imagens do que o experimento apresentado pelo Quadro 7, porém mais do que o experimento apresentado no Quadro 8. Além disso, acredita-se que, nos dois experimentos apresentados a seguir, a quantidade de faces falsas detectadas foi relativamente baixa em relação ao total de imagens detectadas e de faces verdadeiras, tornando assim os resultados mais atraentes.

O primeiro deles foi realizado utilizando o classificador pré-treinado com o arquivo *Haar-cascade frontal face alternative 2* com os parâmetros:

Quadro 9 – Parâmetros do algoritmo de detecção facial

XML: Haar-cascade frontal face alternative 2		
scaleFactor	minNeighbors	minSize
1.05	5	30

Fonte: Kelvin Salton do Prado, 2017.

Este, por sua vez, detectou 1.754 imagens ao total, sendo 397 faces falsas (cerca de 22.63%) e 1.357 faces verdadeiras (cerca de 77.37%). Acredita-se que este resultado é mais relevante que os resultados anteriores, pois apesar de a taxa de faces verdadeiras ser mais baixa que a mesma taxa apresentada no experimento anterior, a quantidade total de faces detectadas é superior, se comparada ao total de faces detectadas no mesmo experimento.

Cabe salientar que, quando há variação de qualquer um dos três parâmetros utilizados, nota-se a diferença nos resultados finais. Reduzir quaisquer um dos parâmetros resulta em um número maior de detecções. Por outro lado, isso tende a aumentar o número de imagens faciais falsas detectadas, como é possível visualizar na Tabela 1 apresentada adiante nessa seção.

O segundo experimento considerado relevante foi realizado utilizando o classificador pré-treinado com o arquivo *Haar-cascade frontal face alternative 2* e os seguintes parâmetros:

Quadro 10 – Parâmetros do algoritmo de detecção facial

XML: Haar-cascade frontal face alternative 2		
scaleFactor	minNeighbors	minSize
1.2	5	10

Fonte: Kelvin Salton do Prado, 2017.

Nesse caso, foram detectadas 634 imagens, sendo apenas 9 faces falsas (aproximadamente 1.42% do total) e 625 faces verdadeiras (cerca de 98.58% do total). Neste experimento é possível notar claramente que a taxa de faces verdadeiras detectadas é superior aos outros experimentos, chegando quase a 100%. Por outro lado, o total de faces detectadas é consideravelmente mais baixo que o experimento anterior que, por sua vez, detectou um total de 1.754 imagens.

Concluindo a análise dos dois experimentos considerados mais relevantes, é possível resumir que o primeiro experimento obteve uma taxa de faces verdadeiras inferior ao segundo, porém a quantidade de imagens faciais detectadas foi superior, ou seja, mesmo tendo uma taxa de faces falsas maior foi possível obter mais resultados, se comparado ao outro experimento considerado relevante. O segundo experimento considerado relevante teve uma taxa de faces falsas consideravelmente menor que o anterior, porém também teve uma menor quantidade de imagens faciais detectadas. Apesar disso, acredita-se que o segundo experimento apresentou um resultado melhor, com relação ao objetivo do presente trabalho, pois teve uma taxa de imagens faciais falsas detectadas menor que o experimento anterior, o que pode ser um ponto positivo, tendo em consideração a baixa qualidade dos vídeos obtidos para os testes.

Como as possíveis combinações entre os 3 parâmetros são diversas, é inviável a realização de experimentos com todas elas neste trabalho, principalmente devido ao tempo e ao escopo da pesquisa. Por este motivo, optou-se por realizar os testes com algumas combinações de valores que aparentemente se mostraram mais representativas. Neste contexto, a Tabela 1 apresenta um resumo dos testes realizados na etapa de detecção facial, com o objetivo de identificar o conjunto de parâmetros que se adaptasse ao ambiente onde os experimentos foram realizados. Nessa tabela, a linha correspondente aos parâmetros selecionados está destacada com a cor de fundo azul e o texto em negrito.

Tabela 1 – Resumo dos testes realizados para a seleção dos parâmetros da etapa de detecção facial

Cascade Classifier (XML)	Video	Scale Factor	Min. Neighbors	Min. Size	Imagens Detectadas	Imagens Faciais	Imagens Não Faciais
Haar-cascade Frontal Face Default	2	1.05	5	30	3588	1303	2285
LBP-cascade Frontal Face	2	1.05	5	30	1502	460	1042
Haar-cascade Frontal Face alt2	2	1.05	5	30	1754	1357	397
Haar-cascade Frontal Face alt tree	2	1.05	5	30	839	760	79
Haar-cascade Frontal Face Default	2	1.1	5	30	1297	618	679
LBP-cascade Frontal Face	2	1.1	5	30	397	240	157
Haar-cascade Frontal Face alt2	2	1.1	5	30	803	708	95
Haar-cascade Frontal Face alt tree	2	1.1	5	30	287	272	15
Haar-cascade Frontal Face Default	2	1.2	5	10	710	519	191
LBP-cascade Frontal Face	2	1.2	5	10	215	194	21
Haar-cascade Frontal Face alt2	2	1.2	5	10	634	625	9
Haar-cascade Frontal Face alt tree	2	1.2	5	10	318	313	5

Fonte: Kelvin Salton do Prado, 2017.

Assim, através da análise dos resultados, foi possível selecionar um conjunto de parâmetros que se mostraram mais atraentes de acordo com os resultados obtidos. Neste caso, acredita-se que o conjunto de parâmetros apresentados no Quadro 10 se mostraram mais significantes para o problema em questão. Como apresentado anteriormente, a

utilização dos parâmetros demonstrados nesse quadro resultou em um número razoável de faces detectadas e uma baixa taxa de faces falsas, por isso a escolha destes parâmetros.

Com o objetivo de exemplificar as imagens faciais detectadas nos vídeos, a Figura 33 apresenta uma amostra das imagens faciais detectadas pelo algoritmo utilizando os parâmetros apresentados no Quadro 10. Como é possível visualizar na figura, a qualidade das imagens obtidas não é ideal para o processo de reconhecimento facial, pois possui principalmente variação de pose e baixa nitidez.

Figura 33 – Imagens faciais detectadas pelo algoritmo.

Fonte: Kelvin Salton do Prado, 2017.

Além disto, o algoritmo também aponta algumas imagens como sendo faces, mas que na verdade não correspondem a imagens faciais, chamadas neste trabalho de faces falsas ou imagens não faciais, as quais podem ser visualizadas no exemplo apresentado na Figura 34. Essas faces falsas, por sua vez, geram um aumento desnecessário no custo computacional, podendo até mesmo prejudicar os resultados do reconhecimento facial. Como as imagens detectadas e extraídas dos vídeos foram classificadas manualmente, este problema é minimizado nos experimentos, pois é possível mensurar a quantidade de imagens não faciais e até mesmo ignorá-las durante os experimentos envolvendo o reconhecimento facial. A mensuração das imagens faciais e não faciais detectadas em cada vídeo é explicada com mais detalhes na Seção 4.3.3.

Figura 34 – Imagens não faciais detectadas pelo algoritmo.

Fonte: Kelvin Salton do Prado, 2017.

Tendo selecionado os parâmetros utilizados na etapa de detecção facial foi então realizada a detecção e extração das imagens faciais em todos os vídeos de teste. Na seção seguinte serão apresentados os resultados obtidos na etapa de detecção facial em cada vídeo.

4.3.3 Resultados da detecção facial

A seguir serão apresentados os resultados obtidos durante a etapa de detecção facial em cada um dos vídeos utilizados nos testes. A Tabela 2 apresenta um resumo com informações das imagens detectadas em cada vídeo. Tais resultados foram obtidos por meio do uso do algoritmo Viola-Jones, utilizando os parâmetros descritos no Quadro 10 da seção anterior.

Tabela 2 – Resumo dos resultados da operação de detecção facial em cada vídeo

Vídeo	Tempo total	Quant. de frames	Total de imagens detectadas	Total de imagens faciais conhecidas	Porcentagem de imagens faciais conhecidas	Total de imagens não faciais	Total de imagens não conhecidas
1	00:59:57	89.933	556	265	47.66 %	217	74
2	00:01:42	3.076	634	577	91.01 %	8	49
3	00:36:21	65.390	273	123	45.05 %	71	79
4	00:59:56	89.915	19	0	00.00 %	19	0
5	00:59:56	89.911	5774	738	12.78 %	4735	301
6	00:57:50	86.757	307	55	17.91 %	250	2
7	00:09:45	14.641	137	2	01.46 %	135	0
8	00:06:14	11.229	365	0	00.00 %	365	0
9	01:07:01	100.539	903	333	36.88 %	404	166
10	00:59:55	89.893	2137	735	34.39 %	407	995
11	00:59:56	89.907	788	192	24.36 %	530	66
12	00:59:56	89.913	1232	340	27.60 %	511	381

Fonte: Kelvin Salton do Prado, 2017.

Cabe esclarecer que a classificação das imagens foi realizada de forma manual, sendo que as imagens classificadas como “imagens faciais conhecidas” correspondem às imagens da face dos sujeitos encontrados na base de treinamento. As imagens classificadas como “não faciais”, por sua vez, correspondem às imagens apontadas como sendo faces pelo algoritmo de detecção facial, mas que na verdade não correspondiam a uma face. Por fim, as imagens classificadas como “não conhecidas” correspondem aos seguintes casos:

- Imagens que correspondem a imagens faciais, porém nas quais não é possível identificar a qual sujeito ela pertence (devido, principalmente, à sua baixa qualidade).
- Imagens faciais que não pertencem aos participantes da pesquisa (base de treinamento). Como trata-se de um ambiente semicontrolado, outros estudantes e até mesmo funcionários da academia podem ter a imagem capturada pelas câmeras e, neste caso, estas imagens serão classificadas como “não faciais ou não conhecidas”.

Por meio da Tabela 2 é possível notar que o vídeo 2 teve uma taxa maior de imagens faciais conhecidas. Isto ocorre porque o vídeo 2 não corresponde a um vídeo obtido das câmeras de segurança da academia, e sim, a um vídeo gravado com um dispositivo móvel durante uma apresentação dos alunos. Ainda assim, optou-se por manter este vídeo nos testes, pois ele pode ser utilizado na comparação dos algoritmos (objetivo principal deste trabalho) e, como a qualidade da câmera do dispositivo é melhor do que a qualidade das câmeras de segurança da academia, pode servir como um exemplo de que a utilização de câmeras com melhor resolução poderiam melhorar os resultados obtidos durante o processo de reconhecimento facial.

Complementando a Tabela 2 mostrada anteriormente, apresenta-se a seguir a Tabela 3, a qual exhibe os resultados relacionados à quantidade de participantes detectados em cada vídeo. A coluna denominada “Número da Sala” diz respeito ao número da sala correspondente ao vídeo (uma legenda equivalente ao número de cada sala pode ser vista na Figura 35 após a tabela), a coluna “Participantes no Vídeo” corresponde a quantidade dos participantes desta pesquisa que aparecem no vídeo e, por fim, a coluna denominada “Participantes Detectados” reflete a quantidade dos participantes que foram detectados pelo algoritmo de detecção facial em cada vídeo.

Tabela 3 – Resumo da quantidade de participantes detectados em cada vídeo.

Número do Vídeo	Número da Sala	Participantes no Vídeo	Participantes Detectados
1	1	7	5
2	2.a	8	4
3	1	6	4
4	3	1	0
5	2.b	6	6
6	1	7	2
7	2.b	1	1
8	2.b	1	0
9	2.b	1	1
10	2.b	5	5
11	2.b	7	6
12	2.b	7	7

Fonte: Kelvin Salton do Prado, 2017.

Cabe esclarecer que estes resultados foram contabilizados manualmente e independente do número de vezes que o participante apareceu no vídeo, ou seja, mesmo que o participante tenha aparecido apenas por alguns instantes (em alguns *frames*), ele foi

contabilizado. Ainda, explica-se que as salas denominadas 2.a e 2.b correspondem à mesma sala, porém os vídeos foram gravados com ângulos e câmeras diferentes.

Figura 35 – Exemplo das 3 salas com os respectivos números.

Fonte: Kelvin Salton do Prado, 2017.

Como é possível visualizar na Tabela 3 apresentada anteriormente, nem todos os participantes que aparecem no vídeo tem sua face detectada pelo algoritmo, o que prejudica consequentemente o reconhecimento automático dos alunos e instrutores no ambiente real da academia. Acredita-se que isso acontece por diversos fatores, dentre eles a baixa nitidez das imagens, a variação de pose e a variação de iluminação.

Esclarece-se ainda que, apesar da detecção facial ser uma parte essencial do processo de reconhecimento facial, neste trabalho não temos esta etapa como foco principal pois, como é possível notar, o processo de detecção de face, envolvendo algoritmos e parâmetros, é uma área de estudo relativamente grande e que merece estudos específicos. Nas próximas seções serão apresentados os experimentos realizados e resultados encontrados na etapa de reconhecimento facial.

4.4 Pré-processamento dos dados

O pré-processamento é uma etapa importante e obrigatória, pois os métodos de reconhecimento facial utilizados necessitam que as imagens tenham os mesmos padrões de cor e dimensão. Assim, precedentemente ao reconhecimento facial, é necessária a realização do pré-processamento das imagens extraídas dos vídeos, para que elas estejam todas com um mesmo padrão de dimensão e espaço de cores. Este processamento deve ser realizado tanto para as imagens de treinamento quanto para as imagens de teste.

O pré-processamento pode ser basicamente dividido em duas etapas, sendo elas:

1. **Conversão para escala de cinza:** como alguns dos algoritmos de reconhecimento facial (e.g. eigenfaces) trabalham apenas com imagens em escala de cinza, é necessário que primeiramente todas as imagens sejam convertidas para esta escala. A própria biblioteca OpenCV disponibiliza uma função chamada “*cvtColor*” que, por sua vez, pode ser utilizada com o parâmetro “*COLOR_RGB2GRAY*” para realizar esta conversão. A função basicamente converte os valores vermelho, verde e azul (RGB) correspondentes à intensidade de cada pixel i para um valor em escala de cinza (0 a 255) de acordo com a seguinte fórmula (OPENCV, 2000):

$$Y_i = 0.299 * R_i + 0.587 * G_i + 0.114 * B_i \quad (34)$$

Assim, a função calcula o valor da intensidade do novo pixel em escala de cinza adicionando 30% do vermelho, 59% do verde e 11% do azul do pixel original.

2. **Redimensionamento da imagem:** assim como a necessidade de se utilizar imagens em escala de cinza, alguns algoritmos (e.g. fisherfaces) necessitam que as imagens de treinamento e de testes tenham uma mesma dimensão (largura e altura). Como a dimensão das imagens faciais obtidas dos vídeos varia, principalmente de acordo com a distância dos sujeitos em relação à câmera, é necessário realizar o redimensionamento das mesmas. Para isto, foi utilizado um método chamado *resize*, também disponibilizado pela biblioteca OpenCV, o qual é responsável por realizar o redimensionamento com base no método de interpolação passado como parâmetro. Como já mencionado na Seção 3.8, neste trabalho foi utilizado o método de interpolação bicúbica para a realização do redimensionamento das imagens. Uma breve comparação sobre os métodos de interpolação pode ser encontrada na Seção 4.4.1. Além disso, foi utilizada

uma dimensão padrão de 100×100 em todos os experimentos. Cabe esclarecer que, como a dimensão das imagens faciais detectadas varia de 21×21 a 87×87 *pixels*, assim como apresentado no Quadro 6 da Seção 4.2.4, optou-se por utilizar a dimensão padrão de 100×100 *pixels*, pois desta maneira não serão removidos *pixels* das imagens de teste durante o redimensionamento e a dimensão final ficará, em geral, próxima da dimensão da imagem original.

Na seção seguinte será feita uma breve comparação sobre os métodos de interpolação utilizados no redimensionamento das imagens, com o intuito de esclarecer a escolha da interpolação bicúbica.

4.4.1 Redimensionamento das Imagens

Conforme mencionado, um dos parâmetros utilizados pelo método *resize* da OpenCV é a *flag interpolation*, que representa qual algoritmo de interpolação o método irá utilizar. A OpenCV disponibiliza 5 algoritmos de interpolação, a saber:

- (b) INTER_AREA : reamostragem utilizando a relação de área de *pixels*. Quando a imagem é ampliada, este método é semelhante ao método do vizinho mais próximo.
- (c) INTER_CUBIC : interpolação bicúbica (*bicubic interpolation*) sobre uma vizinhança de 4×4 pixels.
- (d) INTER_LANCZOS4 : interpolação Lanczos (*Lanczos interpolation*) sobre uma vizinhança de 8×8 pixels.
- (e) INTER_LINEAR : interpolação bilinear (*bilinear interpolation*).
- (f) INTER_NEAREST : interpolação do vizinho mais próximo (*nearest-neighbor interpolation*).

Diferentes resultados são obtidos aplicando-se os diferentes métodos de interpolação, assim como pode-se visualizar na Figura 36. Nela, são apresentados os resultados do redimensionamento de uma imagem de tamanho 50×50 (a) para uma dimensão de 100×100 utilizando cada um dos algoritmos de interpolação disponibilizados pela OpenCV. Como pode-se observar, os métodos utilizados pelas *flags* INTER_AREA e INTER_NEAREST resultam em uma imagem onde é possível visualizar o efeito dos *pixels*, principalmente nas bordas, como por exemplo no chapéu. Já os métodos utilizados pelas *flags* INTER_CUBIC

e INTER_LANCZOS4 resultam em uma imagem mais suavizada e aparentemente mais semelhante à original, enquanto a *flag* INTER_LINEAR, apesar de suavizar as bordas, gera uma imagem um pouco mais “desfocada”, se comparada as demais.

Figura 36 – Resultado de cada método de interpolação.

Fonte: Kelvin Salton do Prado, 2017.

Com o intuito de identificar o método que melhor se adapte aos objetivos deste trabalho, foram realizados alguns testes variando o algoritmo de interpolação utilizado e a dimensão da imagem padrão aplicados aos algoritmos de reconhecimento facial. Os resultados dos testes podem ser vistos no Apêndice F, onde a coluna denominada “FACES Reconhecidas” corresponde à quantidade de faces corretamente reconhecidas pelo algoritmo, a coluna “FACES Não Reconhecidas” corresponde às imagens faciais que foram incorretamente reconhecidas pelo algoritmo e, por fim, a coluna “Não são Faces” corresponde às imagens apontadas como sendo faces pelo algoritmo de detecção facial, mas que na verdade não eram faces, e pelas imagens faciais não pertencentes aos sujeitos da base de treinamento.

Por meio dos testes apresentados no Apêndice F foi possível concluir que, de forma geral, o método do vizinho mais próximo teve um resultado inferior aos demais métodos. Porém não foi possível identificar um padrão que indique qual dos métodos de interpolação teve o melhor resultado no ambiente proposto, pois os resultados foram de certa forma semelhantes e variaram bastante de acordo com as bases de treinamento e dimensões utilizadas.

Sendo assim, optou-se por utilizar o método de interpolação bicúbica, pois seu resultado nos testes foi, de certa forma, semelhante aos demais e, devido ao seu desempenho satisfatório, ele é um método utilizado em diversas aplicações na área de processamento de imagens, assim como explicado na Seção 3.8.

4.5 *Reconhecimento facial*

Nas seções seguintes serão apresentados os experimentos realizados, bem como os resultados obtidos em cada experimento. Cabe primeiramente apresentar os parâmetros utilizados por cada algoritmo disponibilizado pela biblioteca OpenCV, assim como uma breve descrição e os valores padrão de cada parâmetro, os quais são apresentados na Seção 4.5.1 a seguir.

4.5.1 Descrição dos parâmetros

Para melhor entender como os experimentos foram conduzidos, e assim compreender o trabalho como um todo, a seguir é apresentada uma lista com todos os parâmetros de todos os algoritmos utilizados neste trabalho, bem como uma descrição sobre a utilização de cada parâmetro e o seu valor padrão utilizado pela biblioteca. Ainda, a descrição dos parâmetros apresentados a seguir, bem como o valor padrão de cada parâmetro, foram obtidos da documentação oficial da OpenCV (OPENCV, 2000).

- **Eigenfaces**

- **num_components**: Representa o número de componentes mantidos para a Análise de Componentes Principais (PCA). Não existe uma regra de quantos componentes devem ser mantidos para uma boa capacidade de reconstrução. Porém, de acordo com a documentação da OpenCV, manter até 80 componentes deve ser, na maior parte dos casos, suficiente. O valor padrão deste parâmetro é 0, o que significa que serão utilizados todos os componentes (OPENCV, 2000).
- **threshold**: Limiar aplicado na classificação. O valor padrão deste parâmetro é DBL_MAX, o que significa que o algoritmo irá utilizar o valor máximo para o *threshold* e, neste caso, todos os resultados serão retornados pois estarão abaixo do limiar. Cabe esclarecer que o limiar utilizado nos experimentos não está

relacionado a este parâmetro e, apesar de ter o mesmo objetivo, ele é utilizado apenas após a classificação para medir os resultados de acordo com a confiança retornada de cada algoritmo.

• LBPH

- **radius**: O raio utilizado para construir o LBP circular. A OpenCV utiliza o valor padrão 1.
- **neighbors**: O número de pontos de amostra utilizados para construir o LBP circular. De acordo com a documentação da OpenCV, um valor apropriado para este parâmetro é utilizar 8 pontos de amostra, porém cabe esclarecer que quanto mais pontos de amostra forem incluídos, maior será o custo computacional (OPENCV, 2000). Utiliza o valor padrão 8.
- **grid_x**: O número de células, na direção horizontal, que serão utilizadas para construir o histograma. De acordo com a documentação da OpenCV (2000) 8 é um valor padrão utilizado em diversas publicações, porém é possível encontrar trabalhos onde os experimentos são realizados variando a quantidade de células (e.g. (AHONEN; HADID; PIETIKÄINEN, 2004)). Cabe esclarecer ainda que quanto maior o número de células, mais fina é a grade e, consequentemente, maior é a dimensionalidade do vetor de características resultante. Utiliza o valor padrão 8.
- **grid_y**: O número de células, na direção vertical, que serão utilizadas para construir o histograma. De acordo com a documentação da OpenCV (2000) 8 é um valor padrão utilizado em diversas publicações, porém é possível encontrar trabalhos onde os experimentos são realizados variando a quantidade de células (e.g. (AHONEN; HADID; PIETIKÄINEN, 2004)). Cabe esclarecer ainda que quanto maior o número de células mais fina é a grade e, consequentemente, maior é a dimensionalidade do vetor de características resultante. Utiliza o valor padrão 8.
- **threshold**: Limiar aplicado na classificação. Assim como no Eigenfaces, o valor padrão desse parâmetro é DBL_MAX, que corresponde ao valor máximo para o *threshold*. Como explicado anteriormente, nos parâmetros do Eigenfaces, este parâmetro não é utilizado nos experimentos.

- **Fisherfaces**

- **num_components**: Número de componentes mantidos para a LDA com o critério de Fisherfaces. É útil manter todos os componentes, ou seja, o número de classes c (sujeitos que deseja-se reconhecer). Se este valor for definido para menor que zero, igual a zero ou maior que o número de classes menos 1 ($c-1$), ele será automaticamente definido para o número correto de classes ($c-1$) (OPENCV, 2000). Este parâmetro utiliza o valor padrão 0, ou seja, irá utilizar o número de classes ($c-1$).
- **threshold**: Limiar aplicado na classificação. Assim como no Eigenfaces, o valor padrão desse parâmetro é DBL_MAX, que corresponde ao valor máximo para o *threshold*. Como explicado anteriormente, nos parâmetros do Eigenfaces, este parâmetro não é utilizado nos experimentos.

- **SIFT**

- **nfeatures**: Número de melhores características para manter. As características são classificadas por suas pontuações (mensuradas no SIFT como o contraste local). Valor padrão: 0.
- **nOctaveLayers**: Número de camadas em cada oitava. 3 é o valor utilizado na publicação original de David G. Lowe (OPENCV, 2000). O número de oitavas é calculado automaticamente a partir da dimensão da imagem. Valor padrão: 3.
- **contrastThreshold**: Limiar de contraste utilizado para filtrar as características fracas em regiões semi-uniformes (de baixo contraste). Quanto maior o limiar, menos características são produzidas pelo detector. Valor padrão: 0.04.
- **edgeThreshold**: Limiar utilizado para filtrar recursos de borda. Diferente do *contrastThreshold*, quanto maior este limiar menos características são filtradas (mais características são retidas). Valor padrão: 10.
- **sigma**: Sigma da Gaussiana, aplicado à imagem de entrada na oitava #0. Valor padrão: 1.6.

- **SURF**

- **hessianThreshold**: Limiar utilizado para o detector de pontos-chave do SURF. Apenas características onde a Hessiana é maior que este limiar serão retidas. Portanto, quanto maior este valor, menos pontos-chave serão obtidos. Ainda, de acordo com a documentação da OpenCV (2000), um valor geralmente utilizado

é entre 300 e 500, dependendo do contraste da imagem. A biblioteca utiliza o valor padrão 100.

- **nOctaves**: Número de oitavas de pirâmide que o detector de ponto chave utilizará. Valor padrão: 4.
- **nOctaveLayers**: Número de camadas de oitavas dentro de cada oitava. Valor padrão: 3.
- **extended**: Indicador de descritor estendido (true - utiliza descritores de 128 elementos estendidos; false - utiliza descritores de 64 elementos). Valor padrão: False.
- **upright**: Indicador de características para cima e a direita ou para rotação (true - não calcula a orientação das características; false - calcula a orientação). Valor padrão: False.

Cabe salientar que a descrição dos parâmetros, apresentados na lista anterior, foi obtida da documentação oficial da biblioteca OpenCV (2000), e diz respeito aos parâmetros que podem ser utilizados em cada um dos algoritmos disponibilizados pela mesma. Ainda, cabe explicar que neste trabalho foram utilizados os parâmetros padrão de cada algoritmo e que, de forma geral, os mesmos foram definidos como padrão com base nas especificações apresentadas nos artigos originais de cada método.

Além dos parâmetros descritos acima, o algoritmo desenvolvido neste trabalho também utiliza um parâmetro chamado de *threshold* ou limiar, que é utilizado para definir, de acordo com a confiança, se a face foi corretamente reconhecida ou não. O próprio algoritmo de reconhecimento facial da OpenCV retorna um valor de confiança que representa o quanto aquela face corresponde ao sujeito que o algoritmo classificou, sendo que quanto mais próxima de zero maior é a confiança do algoritmo. O parâmetro *threshold* foi definido de acordo com alguns testes realizados em um ambiente controlado com uma base de imagens criada pelo autor para este fim, porém acredita-se que o mesmo deva ser ajustado de acordo com as necessidades e com os resultados obtidos nos testes com as imagens reais da academia. Tanto o parâmetro *threshold* quanto o valor da confiança serão explicados detalhadamente nas Seções 4.5.2 e 4.5.3.

4.5.2 Resultados principais

Com as imagens previamente detectadas e classificadas manualmente é possível então realizar a etapa de reconhecimento facial e medir os resultados dos algoritmos, comparando-os às imagens previamente classificadas. Sendo assim, nesta seção serão apresentados os experimentos realizados, bem como os resultados obtidos em cada experimento.

Cabe primeiramente recordar que, além de realizar a classificação do sujeito, os algoritmos de reconhecimento facial também calculam um valor chamado de confiança, que no caso dos métodos Eigenfaces, Fisherfaces e LBPH corresponde à medida da distância entre a imagem facial submetida e a imagem facial classificada, e no caso dos métodos SIFT e SURF é calculada de acordo com a quantidade de pontos-chave combinados entre a imagem facial submetida e a imagem facial classificada, em relação ao total de pontos encontrados na imagem submetida, conforme a seguinte equação:

$$\text{Confiança} = 100 - \frac{\text{Pontos que Combinam} * 100}{\text{Total de Pontos Característicos}} \quad (35)$$

Note que em ambos os casos, utilizando a medida de distância ou os pontos característicos, quanto mais próximo de zero for a confiança maior é a chance da classificação estar correta. Apesar de o termo “confiança” parecer erroneamente empregado neste contexto, ele foi adotado pois é o mesmo termo utilizado pela biblioteca OpenCV⁷ nas funções de classificação.

Com isso, é possível utilizar esta medida de confiança para julgar se a imagem facial foi corretamente reconhecida ou não, sem a necessidade, neste caso, de classificar previamente e de forma manual as imagens coletadas, o que ocorre geralmente em sistemas de reconhecimento facial comerciais, onde não é viável a classificação manual de todas as imagens detectadas. No presente caso, contudo, torna-se viável a classificação manual, pois os experimentos foram realizados com um número limitado de imagens e vídeos.

Tendo compreendido isto, cabe esclarecer que os experimentos serão divididos em duas partes, sendo que a primeira parte não faz uso do valor de confiança, pois tem como objetivo comparar os algoritmos de reconhecimento facial entre si, utilizando para tal as imagens previamente classificadas. Já a segunda parte dos experimentos faz uso do valor da confiança, pois tem como objetivo analisar a possibilidade de utilizar os algoritmos de

⁷ A OpenCV utiliza o termo em inglês “*confidence*”.

reconhecimento facial para realizar a detecção de presença dos alunos em sala de aula e, neste caso, é necessário supor que se o sistema fosse utilizado no ambiente real não seria possível classificar manualmente as imagens faciais para comparar com os resultados dos algoritmos.

Antes de apresentar os experimentos e os resultados obtidos, cabe esclarecer que, como mencionado na Seção 4.2.4, 3 dos 12 vídeos coletados foram removidos dos experimentos, por não apresentarem resultados satisfatórios na etapa de detecção facial, como é possível visualizar na Tabela 2 da Seção 4.3.3. Como os vídeos foram nomeados de acordo com a ordem de coleta, foram removidos dos testes os vídeos 4, 7 e 8, e, por este motivo, os resultados do reconhecimento facial em tais vídeos não serão incluídos nas tabelas apresentadas a seguir.

Além disso, nos experimentos apresentados a seguir não foram utilizadas as imagens não faciais (apontadas como sendo faces pelo algoritmo de detecção facial, mas que na verdade não correspondiam a uma face) e as que não pertencem aos sujeitos da base de treinamento. Assim, os experimentos foram realizados utilizando apenas imagens faciais de sujeitos que realmente se encontram nas bases de treinamento. Ainda, as imagens não utilizadas nos experimentos, descritas neste parágrafo, foram classificadas de forma manual, desta forma é possível remover imagens irrelevantes que podem ser consideradas “ruídos” nos dados, buscando resultados mais expressivos⁸ em relação ao reconhecimento facial.

Esse procedimento, contudo, não traz necessariamente uma melhora nos resultados de precisão. As imagens consideradas “ruídos” são geralmente classificadas como verdadeiro negativo (VN) e, neste caso, não alteram o resultado da precisão. Tais imagens irão afetar a precisão negativamente apenas quando elas forem classificadas como falso positivo (FP), ou seja, quando o algoritmo classificar um “ruído” como pertencente a um sujeito da base de treinamento, com uma confiança aceitável (abaixo do limiar).

Explica-se ainda que, neste contexto, como não se está utilizando um valor de limiar, os resultados são binários, ou seja, o algoritmo irá apresentar um resultado correto ou incorreto. Sendo assim, é possível considerar que os resultados corretos são verdadeiros positivos e os incorretos são falsos positivos, e, a partir disso, utilizar a Equação 30 apresentada na Seção 3.7 para calcular a precisão.

⁸ O termo “expressivos” é utilizado aqui com a intenção de explicar que, ao remover os “ruídos”, é possível demonstrar melhor os resultados dos algoritmos de acordo com os objetivos do trabalho.

Cabe ainda colocar que, na fase de pré-processamento, as imagens foram convertidas para escala de cinza e foram redimensionadas para um tamanho padrão, neste caso 100×100 *pixels*, utilizando a interpolação bicúbica, apresentada nas Seções 3.8 e 4.4.1.

Além disso, o comitê de máquinas, cujos resultados são apresentados nas Tabelas 4 a 9, utiliza os 5 algoritmos como componentes e faz uso do esquema de voto majoritário para combinar os resultados. Como os resultados apresentados nestas tabelas são de testes iniciais, não se tem, neste primeiro momento, parâmetros suficientes para utilizar o sistema de voto ponderado, sendo que, neste esquema de votação, o ideal seria atribuir maiores pesos para os algoritmos que tem melhores resultados individualmente.

A seguir, a Tabela 4 apresenta os resultados da precisão de todos os algoritmos treinados utilizando a primeira base de treinamento descrita na Seção 4.2.3.

Tabela 4 – Precisão dos algoritmos utilizando a primeira base de treinamento

	Eigenfaces	LBPH	Fisherfaces	SIFT	SURF	Comitê de Máquinas
Vídeo 1	00.00 %	01.89 %	00.00 %	04.15 %	10.57 %	01.13 %
Vídeo 2	01.04 %	01.73 %	01.39 %	02.43 %	07.80 %	01.21 %
Vídeo 3	00.00 %	13.01 %	00.00 %	01.63 %	01.63 %	00.00 %
Vídeo 5	05.83 %	17.75 %	05.28 %	02.98 %	02.98 %	06.10 %
Vídeo 6	00.00 %	03.64 %	00.00 %	01.82 %	05.45 %	00.00 %
Vídeo 9	00.00 %	00.00 %	00.00 %	02.70 %	02.10 %	00.00 %
Vídeo 10	00.14 %	05.85 %	00.41 %	04.49 %	02.18 %	00.54 %
Vídeo 11	19.27 %	15.10 %	18.23 %	02.60 %	04.17 %	19.27 %
Vídeo 12	02.06 %	04.71 %	02.06 %	05.88 %	05.88 %	01.47 %
Mínima	00.00 %	00.00 %	00.00 %	01.63 %	01.63 %	00.00 %
Máxima	19.27 %	17.75 %	18.23 %	05.88 %	10.57 %	19.27 %
Média	03.15 %	07.07 %	03.04 %	03.19 %	04.75 %	03.30 %

Fonte: Kelvin Salton do Prado, 2017.

A Figura 37 apresenta um gráfico com a precisão média (representada na última linha da tabela 4) de cada algoritmo, com base em todos os vídeos de teste, utilizando a primeira base de treinamento, ou seja, a base formada apenas pelas imagens da ficha de cadastro dos participantes.

Figura 37 – Precisão média de cada algoritmo utilizando a primeira base de treinamento.

Fonte: Kelvin Salton do Prado, 2017.

A seguir, a Tabela 5 apresenta os resultados da precisão de todos os algoritmos treinados utilizando a segunda base de treinamento descrita na Seção 4.2.3, a saber, a base formada pelas imagens do cadastro e pelas imagens geradas por meio da utilização do primeiro nível da Wavelet *db2*.

Tabela 5 – Precisão dos algoritmos utilizando a segunda base de treinamento

	Eigenfaces	LBPH	Fisherfaces	SIFT	SURF	Comitê de Máquinas
Vídeo 1	00.00 %	02.64 %	00.38 %	01.89 %	10.94 %	01.13 %
Vídeo 2	01.21 %	01.91 %	01.91 %	04.33 %	07.45 %	01.21 %
Vídeo 3	00.00 %	08.94 %	00.00 %	01.63 %	00.81 %	00.00 %
Vídeo 5	05.01 %	17.21 %	04.07 %	02.57 %	02.71 %	04.20 %
Vídeo 6	00.00 %	00.00 %	00.00 %	00.00 %	03.64 %	00.00 %
Vídeo 9	00.00 %	00.00 %	00.00 %	03.60 %	03.00 %	00.00 %
Vídeo 10	00.14 %	07.21 %	00.41 %	06.39 %	03.27 %	00.54 %
Vídeo 11	16.15 %	11.46 %	17.19 %	03.65 %	03.65 %	16.15 %
Vídeo 12	01.76 %	04.41 %	02.65 %	04.41 %	08.24 %	01.18 %
Mínima	00.00 %	00.00 %	00.00 %	00.00 %	00.81 %	00.00 %
Máxima	16.15 %	17.21 %	17.19 %	06.39 %	10.94 %	16.15 %
Média	02.70 %	05.98 %	02.96 %	03.16 %	04.86 %	02.71 %

Fonte: Kelvin Salton do Prado, 2017.

A Figura 38 apresenta um gráfico com a precisão média de cada algoritmo utilizando a segunda base de treinamento. Cabe esclarecer que a escala desse gráfico, bem como a

dos gráficos apresentados a seguir, foi mantida a mesma da Figura 37, para facilidade de comparação visual.

Figura 38 – Precisão média de cada algoritmo utilizando a segunda base de treinamento.

Fonte: Kelvin Salton do Prado, 2017.

A seguir, a Tabela 6 apresenta os resultados da precisão de todos os algoritmos treinados utilizando a terceira base de treinamento descrita na Seção 4.2.3, a qual utiliza as imagens da ficha de cadastro e as imagens geradas pela utilização do primeiro e segundo nível da transformada Wavelet *db2*.

Tabela 6 – Precisão dos algoritmos utilizando a terceira base de treinamento

	Eigenfaces	LBPH	Fisherfaces	SIFT	SURF	Comitê de Máquinas
Vídeo 1	00.00 %	05.66 %	00.00 %	01.51 %	18.49 %	02.26 %
Vídeo 2	01.56 %	01.56 %	01.21 %	05.37 %	03.99 %	01.39 %
Vídeo 3	00.00 %	05.69 %	00.00 %	02.44 %	00.00 %	00.00 %
Vídeo 5	06.23 %	15.85 %	10.16 %	03.52 %	02.98 %	06.50 %
Vídeo 6	00.00 %	00.00 %	00.00 %	00.00 %	01.82 %	00.00 %
Vídeo 9	00.00 %	00.30 %	00.60 %	07.21 %	06.31 %	00.00 %
Vídeo 10	00.95 %	07.07 %	09.25 %	05.44 %	03.40 %	02.31 %
Vídeo 11	16.67 %	11.46 %	13.54 %	04.17 %	05.21 %	17.19 %
Vídeo 12	02.94 %	04.71 %	09.12 %	06.47 %	07.06 %	05.00 %
Mínima	00.00 %	00.00 %	00.00 %	00.00 %	00.00 %	00.00 %
Máxima	16.67 %	15.85 %	13.54 %	07.21 %	18.49 %	17.19 %
Média	03.15 %	05.81 %	04.88 %	04.01 %	05.47 %	03.85 %

Fonte: Kelvin Salton do Prado, 2017.

A Figura 39 apresenta um gráfico com a precisão média de cada algoritmo utilizando a terceira base de treinamento.

Figura 39 – Precisão média de cada algoritmo utilizando a terceira base de treinamento.

Fonte: Kelvin Salton do Prado, 2017.

A seguir, a Tabela 7 apresenta os resultados da precisão de todos os algoritmos treinados utilizando a quarta base de treinamento (descrita na Seção 4.2.3) que, por sua vez, foi criada a partir das imagens da ficha de cadastro e das imagens geradas pela aplicação da transformada Wavelet discreta de primeiro nível, utilizando como parâmetro a Wavelet *haar*.

Tabela 7 – Precisão dos algoritmos utilizando a quarta base de treinamento

	Eigenfaces	LBPH	Fisherfaces	SIFT	SURF	Comitê de Máquinas
Vídeo 1	00.00 %	01.51 %	00.38 %	02.26 %	14.34 %	01.89 %
Vídeo 2	01.21 %	01.73 %	02.95 %	03.81 %	05.89 %	01.56 %
Vídeo 3	00.00 %	14.63 %	00.00 %	04.07 %	03.25 %	01.63 %
Vídeo 5	05.42 %	18.29 %	04.74 %	05.56 %	02.44 %	04.74 %
Vídeo 6	00.00 %	00.00 %	00.00 %	05.45 %	01.82 %	00.00 %
Vídeo 9	00.00 %	00.00 %	00.00 %	04.80 %	02.40 %	00.00 %
Vídeo 10	00.27 %	04.22 %	01.22 %	03.27 %	03.27 %	00.41 %
Vídeo 11	16.15 %	18.23 %	15.63 %	02.60 %	04.69 %	15.63 %
Vídeo 12	01.76 %	03.82 %	03.53 %	05.88 %	06.47 %	01.76 %
Mínima	00.00 %	00.00 %	00.00 %	02.26 %	01.82 %	00.00 %
Máxima	16.15 %	18.29 %	15.63 %	05.88 %	14.34 %	15.63 %
Média	02.76 %	06.94 %	03.16 %	04.19 %	04.95 %	03.07 %

Fonte: Kelvin Salton do Prado, 2017.

A Figura 40 apresenta um gráfico com a precisão média de cada algoritmo utilizando a quarta base de treinamento.

Figura 40 – Precisão média de cada algoritmo utilizando a quarta base de treinamento.

Fonte: Kelvin Salton do Prado, 2017.

A seguir, a Tabela 8 apresenta os resultados da precisão de todos os algoritmos treinados utilizando a quinta base de treinamento descrita na Seção 4.2.3, a qual utiliza, além das imagens da ficha de cadastro, as imagens geradas pela aplicação do primeiro e segundo nível da transformada Wavelet utilizando o parâmetro *haar*.

Tabela 8 – Precisão dos algoritmos utilizando a quinta base de treinamento

	Eigenfaces	LBPH	Fisherfaces	SIFT	SURF	Comitê de Máquinas
Vídeo 1	00.00 %	01.51 %	00.38 %	01.51 %	08.30 %	01.13 %
Vídeo 2	01.91 %	01.73 %	02.43 %	05.72 %	06.59 %	02.25 %
Vídeo 3	00.00 %	11.38 %	30.08 %	17.07 %	00.81 %	06.50 %
Vídeo 5	08.13 %	18.02 %	13.01 %	04.20 %	02.17 %	10.70 %
Vídeo 6	00.00 %	00.00 %	01.82 %	07.27 %	03.64 %	00.00 %
Vídeo 9	00.00 %	00.00 %	00.30 %	05.11 %	04.50 %	00.00 %
Vídeo 10	03.81 %	05.03 %	08.84 %	04.90 %	02.04 %	04.49 %
Vídeo 11	20.31 %	11.98 %	12.50 %	02.60 %	04.69 %	17.19 %
Vídeo 12	07.65 %	09.41 %	05.88 %	07.06 %	07.65 %	07.65 %
Mínima	00.00 %	00.00 %	00.30 %	01.51 %	00.81 %	00.00 %
Máxima	20.31 %	18.02 %	30.08 %	17.07 %	08.30 %	17.19 %
Média	04.65 %	06.56 %	08.36 %	06.16 %	04.49 %	05.55 %

Fonte: Kelvin Salton do Prado, 2017.

A Figura 41 apresenta um gráfico com a precisão média de cada algoritmo utilizando a quinta base de treinamento.

Figura 41 – Precisão média de cada algoritmo utilizando a quinta base de treinamento.

Fonte: Kelvin Salton do Prado, 2017.

A seguir, a Tabela 9 apresenta os resultados da precisão de todos os algoritmos treinados utilizando a sexta base de treinamento, descrita na Seção 4.2.3, que faz uso de todas as imagens utilizadas nas demais bases, ou seja, as imagens da ficha de cadastro e as imagens geradas pela aplicação do primeiro e segundo nível da Wavelet utilizando os parâmetros *db2* e *haar*.

Tabela 9 – Precisão dos algoritmos utilizando a sexta base de treinamento

	Eigenfaces	LBPH	Fisherfaces	SIFT	SURF	Comitê de Máquinas
Vídeo 1	00.00 %	05.28 %	00.00 %	02.64 %	12.45 %	02.64 %
Vídeo 2	01.73 %	01.56 %	09.19 %	08.32 %	04.68 %	02.25 %
Vídeo 3	00.00 %	04.88 %	00.00 %	13.82 %	00.81 %	00.81 %
Vídeo 5	07.99 %	15.99 %	05.96 %	06.23 %	02.85 %	08.13 %
Vídeo 6	00.00 %	00.00 %	00.00 %	01.82 %	12.73 %	00.00 %
Vídeo 9	00.00 %	00.30 %	00.00 %	09.61 %	06.01 %	01.50 %
Vídeo 10	02.99 %	06.12 %	04.22 %	06.94 %	04.08 %	03.13 %
Vídeo 11	18.75 %	11.46 %	11.46 %	04.69 %	06.25 %	18.75 %
Vídeo 12	07.06 %	09.12 %	15.29 %	06.76 %	07.06 %	10.59 %
Mínima	00.00 %	00.00 %	00.00 %	01.82 %	00.81 %	00.00 %
Máxima	18.75 %	15.99 %	15.29 %	13.82 %	12.73 %	18.75 %
Média	04.28 %	06.08 %	05.12 %	06.76 %	06.32 %	05.31 %

Fonte: Kelvin Salton do Prado, 2017.

A Figura 42 apresenta um gráfico com a precisão média de cada algoritmo utilizando a sexta base de treinamento.

Figura 42 – Precisão média de cada algoritmo utilizando a sexta base de treinamento.

Fonte: Kelvin Salton do Prado, 2017.

Cabe esclarece que, apesar de os parâmetros dos algoritmos serem configuráveis, todos os experimentos, apresentados nas Tabelas 4 a 9 foram realizados utilizando os parâmetros padrão de cada algoritmo, os quais foram introduzidos na Seção 4.5.1. Tais parâmetros foram definidos como padrões pela biblioteca OpenCV seguindo as orientações dos artigos originais de cada método, bem como demais estudos relacionados a estes. Cabe ainda colocar que seria ideal um estudo aprofundado em relação aos parâmetros de cada método, incluindo experimentos que pudessem informar quais parâmetros se comportam melhor no ambiente estudado. Porém, como o tempo e o escopo do trabalho são limitados, optou-se por utilizar os parâmetros padrão da OpenCV, incluindo este estudo como sugestão de trabalhos futuros na Seção 6.1.

Por fim, a Figura 43, apresentada a seguir, representa a precisão média de cada algoritmo utilizando cada uma das bases de treinamento.

Figura 43 – Precisão média de cada algoritmo utilizando cada base de treinamento.

Fonte: Kelvin Salton do Prado, 2017.

Analisando os resultados, é possível notar que as taxas de precisão obtidas por meio dos vídeos de teste 6 e 9 foram relativamente baixas, se comparadas às demais e, em muitos casos, ela foi de 0.00%, ou seja, não reconheceu nem um sujeito. Por outro lado, foi possível obter uma taxa de precisão mais alta com os vídeos 5 e 11, chegando a alcançar uma taxa de até 20.31%.

Como pode-se visualizar nos experimentos apresentados nesta seção, o vídeo 11 apresentou, de forma geral, melhores resultados que os demais. Não foi possível concluir exatamente o motivo dessa ocorrência, porém analisando visualmente o vídeo 11 é possível notar que em diversos momentos os participantes se encontram de frente para a câmera e, por este motivo, acredita-se que a qualidade das imagens faciais capturadas, principalmente em relação à pose, foi ligeiramente melhor, se comparada aos demais vídeos.

Ainda, analisando as médias da precisão em todos os experimentos apresentados anteriormente, é possível concluir que o método LBPH obteve os melhores resultados, seguido do método SURF, e que o método Eigenfaces teve os resultados mais baixos. Além disso, é possível notar que em nenhum momento a média da precisão do comitê de máquinas foi superior a todos os métodos individualmente, o que faz sentido, considerando a colocação de Lima (2004) em que: para que um *ensemble* de classificadores tenha um desempenho melhor do que os componentes independentes, é esperado que os componentes

que o integram tenham um bom desempenho isoladamente e sejam diferentes entre si (HANSEN; SALAMON, 1990 apud LIMA, 2004).

4.5.3 Resultados variando o limiar

Nesta seção serão apresentados alguns experimentos envolvendo o valor de confiança dos algoritmos de reconhecimento facial. Como já justificado na Seção 4.5.2, o valor de confiança pode ser utilizado como um “validador” para a realização do reconhecimento facial em ambientes reais, onde não é viável a classificação ou conferência manual de todas as imagens detectadas.

O ambiente da academia Kungfu-Wushu é um exemplo de um ambiente real onde não é viável a conferência manual de todos os resultados, pois em apenas 1 hora de vídeo é possível gerar mais de 1.000 imagens faciais e, neste caso, seriam necessários vários minutos, ou até mesmo horas, para conferir manualmente todas as imagens. Em casos como este, faz sentido a utilização do valor de confiança para realizar a validação automática do reconhecimento facial, utilizando para tal um valor de limiar (*threshold*) definido pelo usuário.

Como exposto na Seção 4.5.2, o valor de confiança para os métodos Eigenfaces, Fisherfaces e LBPH é calculado pela própria biblioteca OpenCV, de acordo com uma medida de distância, neste caso a medida Euclidiana. Para os métodos SIFT e SURF é calculada com base na Equação 35 que, por sua vez, utiliza os pontos característicos.

É possível notar que em ambos os casos, quanto mais próximo de zero for o valor da confiança maior é a chance do resultado classificado pelo método estar correto. Sendo assim, para o resultado do reconhecimento facial ser considerado correto ele deve estar abaixo do limiar pré definido.

Porém, assim como em outros problemas de classificação, não existe um valor ideal para o parâmetro do limiar, sendo então necessária a realização de experimentos para encontrar um limiar que melhor se adapte ao ambiente estudado.

Levando isto em consideração, faz sentido realizar experimentos com o objetivo de encontrar um valor de limiar que se adapte ao ambiente real desta pesquisa e que auxilie no alcance do objetivo específico deste trabalho, o qual consiste na análise da possibilidade de realizar a detecção de presença automática na academia alvo do estudo.

No âmbito desse experimento, fazendo a utilização do valor de confiança e do limiar, cabe esclarecer que são mensurados os resultados (VP, VN, FP, FN) dos algoritmos de acordo com as medidas e possíveis resultados apresentados na Seção 3.6, mais precisamente nas Figuras 19 e 20, bem como as taxas de verdadeiros positivos (Equação 28) e falsos positivos (Equação 29).

Com a intenção de obter resultados mais representativos, bem como facilitar o entendimento dos resultados, os experimentos apresentados a seguir foram realizados utilizando a quinta e a sexta base de treinamento e o quinto e o décimo primeiro vídeo de teste. Essas bases de treinamento e teste foram selecionadas pois, como pode ser visto nos resultados apresentados na Seção 4.5.2, geraram resultados mais expressivos e, por este motivo, facilitam a visualização e comparação dos resultados. Ainda, cabe justificar que optou-se por não apresentar os demais resultados utilizando as bases de treinamento e de testes restantes pois, como a acurácia resultante nos primeiros experimentos foi baixa (como pode ser visto nas tabelas apresentadas na Seção 4.5.2), os resultados e gráficos apresentados aqui não trariam uma relevância significativa para o entendimento do trabalho.

Antes de apresentar os experimentos, cabe esclarecer que os resultados apresentados a seguir não fazem uso das imagens que não correspondem a faces ou que não pertencem aos participantes da pesquisa. Apesar desses dados influenciarem os resultados de um sistema real de reconhecimento facial, optou-se por não utilizar tais imagens, pois elas iriam alterar significativamente os resultados obtidos, sendo que em alguns vídeos foi detectado um número significativo de imagens não faciais, o que geraria um número maior de verdadeiros negativos e consequentemente poderia alterar os valores da precisão e da taxa de falsos positivos, como explicado anteriormente na Seção 4.5.2.

Cabe ainda observar que as figuras apresentadas a seguir são organizadas da seguinte maneira:

- Gráfico superior esquerdo: apresenta os resultados verdadeiros positivos (azul), falsos positivos (laranja), falsos negativos (amarelo) e verdadeiros negativos (roxo), utilizando a base de treinamento 5 e a base de teste correspondente ao vídeo 5.
- Gráfico superior direito: apresenta os resultados verdadeiros positivos (azul), falsos positivos (laranja), falsos negativos (amarelo) e verdadeiros negativos (roxo), utilizando a base de treinamento 6 e a base de teste correspondente ao vídeo 11.

- Gráfico inferior esquerdo: apresenta a curva ROC (azul), criada utilizando os resultados apresentados no gráfico superior esquerdo (VP, FP, FN e VN), e a curva aleatória (vermelha), utilizada com a finalidade de comparação com os resultados obtidos.
- Gráfico inferior direito: apresenta a curva ROC (azul), criada utilizando os resultados apresentados no gráfico superior direito (VP, FP, FN e VN), e a curva aleatória (vermelha), utilizada com a finalidade de comparação com os resultados obtidos.

A Figura 44 apresenta os resultados da variação do limiar para o método Eigenfaces. Neste experimento os limiares foram variados de 1.000 até 5.000 com um passo de 50. Como é possível visualizar na figura, a curva ROC utilizando a base de treinamento 5 e o vídeo de teste 5 teve um desempenho realmente baixo, pois até mesmo um método de classificação aleatório poderia obter melhores resultados que o Eigenfaces neste cenário. Já na curva ROC utilizando a base de treinamento 6 e o vídeo de teste 11 é possível notar que se tem uma pequena melhora nos resultados utilizando o método Eigenfaces em relação a um método aleatório, pois, apesar de a curva estar inicialmente abaixo da curva aleatória, com o aumento do limiar é possível perceber que ela cruza esta linha, o que significa que a partir de um certo limiar os resultados obtidos pelo método Eigenfaces são melhores que um método de classificação aleatório, ainda assim a diferença é pequena.

Figura 44 – Variação do limiar utilizando o método Eigenfaces.

Fonte: Kelvin Salton do Prado, 2017.

A Figura 45 apresenta os resultados da variação do limiar para o método LBPH. Neste experimento os limiares foram variados de 60 até 200 com um passo de 2. É possível notar claramente que na curva ROC utilizando a base de treinamento 5 e o vídeo 5, tem-se um aumento significativo na taxa de verdadeiros positivos, pois a curva está bem acima da curva aleatória em todos os pontos do gráfico. Sendo assim, pode-se dizer que o método LBPH teve um bom desempenho neste experimento, se comparado aos demais resultados obtidos. Pode-se notar também que, na curva ROC utilizando a base de treinamento 6 e o vídeo 11, houve um aumento na taxa de verdadeiros positivos em relação a um método

de classificação aleatório, porém este aumento não é tão significativo, se comparado ao aumento obtido no experimento utilizando a base de treinamento 5 e o vídeo 5 com o método LBPH.

Figura 45 – Variação do limiar utilizando o método LBPH.

Fonte: Kelvin Salton do Prado, 2017.

A Figura 46 apresenta os resultados da variação do limiar para o método Fisherfaces. Neste experimento os limiares foram variados de 500 até 4.000 com um passo de 50. Na curva ROC utilizando a base de treinamento 5 e o vídeo 5 é possível analisar que, assim como nos resultados obtidos com o Eigenfaces, a taxa de verdadeiros positivos obtida foi baixa, sendo que poderia-se obter um resultado melhor até mesmo utilizando um

método de classificação aleatório. Já na curva ROC utilizando a base de treinamento 6 e o vídeo 11 é possível notar um aumento na taxa de verdadeiros positivos em relação à curva aleatória. É possível notar também que as características das curvas ROC, geradas pelo método Fisherfaces, se assemelham às características das curvas ROC geradas pelo método Eigenfaces. Acredita-se que isto ocorre pois os dois métodos utilizam abordagens semelhantes para a realização do reconhecimento facial.

Figura 46 – Variação do limiar utilizando o método Fisherfaces.

Fonte: Kelvin Salton do Prado, 2017.

A Figura 47 apresenta os resultados da variação do limiar para o método SIFT. Neste experimento os limiares foram variados de 0 até 100 com um passo de 2. Analisando as curvas ROC apresentadas na Figura 47 é possível visualizar que em ambos os experimentos, utilizando as bases de treinamento 5 e 6 e os vídeos de teste 5 e 11, foi possível obter resultados melhores, se comparados a um método de classificação aleatório. Porém, a taxa de verdadeiros positivos obtida no experimento utilizando a base de treinamento 5 e o vídeo 5 com o método SIFT ainda é mais baixa que a obtida no experimento com o método LBPH.

Figura 47 – Variação do limiar utilizando o método SIFT.

A Figura 48 apresenta os resultados da variação do limiar para o método SURF. Neste experimento os limiares foram variados de 0 até 100 com um passo de 2.

Figura 48 – Variação do limiar utilizando o método SURF.

Fonte: Kelvin Salton do Prado, 2017.

Como é possível visualizar na Figura 48, na curva ROC utilizando a base de treinamento 5 e o vídeo de teste 5 a taxa de verdadeiros positivos está acima da curva correspondente a um método aleatório, ou seja, com a utilização do método SURF, neste caso, é possível obter um número de acertos maior do que em uma classificação aleatória. Já na curva ROC utilizando a base de treinamento 6 e o vídeo 11 é possível notar que

tem-se uma variação nos resultados, sendo que em alguns pontos a curva ficou acima da curva aleatória e em outros ficou abaixo. Acredita-se que este resultado não é satisfatório, sendo que espera-se que no mínimo o resultado do método seja igual ou superior a um método aleatório.

Como é possível notar nos gráficos apresentados anteriormente, os valores utilizados como limiares são diferentes para cada método de reconhecimento facial. Isso ocorre pois a confiança é calculada de forma diferente em cada método, assim como explicado no início da Seção 4.5.2.

Por meio das figuras apresentadas acima, considerando os gráficos superiores de cada figura, é possível concluir que o número de verdadeiros positivos cresce de forma proporcional ao número de falsos positivos, ou seja, quanto maior for o limiar utilizado maior é o número de imagens faciais corretamente reconhecidas. Porém, também é maior o número de imagens faciais incorretamente reconhecidas. Além disso, é possível notar que o número de verdadeiros negativos decresce proporcionalmente ao número de falsos negativos de acordo com o aumento do limiar.

Neste contexto, em um sistema de reconhecimento facial real é necessário analisar e selecionar um valor de limiar que se adapte aos dados utilizados, de forma que seja possível reconhecer corretamente um número aceitável de imagens faciais, verdadeiros positivos, e que a quantidade de imagens reconhecidas incorretamente, falsos positivos, não seja tão significativa a ponto de prejudicar a confiabilidade do sistema de reconhecimento facial.

De acordo com os gráficos superiores apresentados nas Figuras 44 a 48, acredita-se que uma boa escolha para os limiares seria aproximadamente nos pontos onde a curva de verdadeiros negativos cruza a de falsos positivos e a de falsos negativos cruza a de verdadeiros positivos, pois desta forma os resultados obtidos estariam de certa forma equilibrados, ou seja, teria-se uma quantidade aceitável de faces corretamente reconhecidas em relação as faces reconhecidas incorretamente (abaixo do limiar). Neste contexto é possível localizar, dentre todas as curvas ROC apresentadas nessa seção, o experimento e o ponto na curva em que a taxa de verdadeiros positivos é maximizada enquanto a taxa de falsos positivos é minimizada, ou seja, o ponto mais próximo ao canto superior esquerdo da curva ROC. Com base nos experimentos realizados, o algoritmo LBPH treinado com a base 5 e testado com o vídeo 5 (Figura 45) apresentou a melhor taxa de verdadeiros positivos (0.6842) em relação a taxa de falsos positivos (0.2760), utilizando o valor de

limiar 120 com o qual foi possível obter um resultado de 91 verdadeiros positivos e 42 falsos negativos.

Cabe ainda esclarecer que estes valores de limiares foram selecionados de acordo com os dados das bases de treinamento e testes deste trabalho, e que em outros casos os mesmos precisam ser analisados de acordo com as bases utilizadas, sendo que é possível selecionar limiares mais baixos, aumentando a confiabilidade dos resultados, quando as imagens de treinamento e testes possuem boa qualidade e são obtidas em cenários semelhantes.

Para complementar a exposição dos resultados obtidos, apresenta-se a seguir as Tabelas 10 e 11, que por sua vez exibem, de forma resumida, os resultados mínimo, máximo e a média da precisão e revocação para cada um dos experimentos apresentados nesta seção.

Tabela 10 – Resumo da precisão e revocação dos experimentos envolvendo a base de treinamento 5 e o vídeo 5

Método	Precisão e Revocação	Mín.	Máx.	Média
Eigenfaces	Precisão	0.00 %	8.02 %	4.03 %
	Revocação	0.00 %	100.00 %	45.55 %
LBPH	Precisão	0.00 %	100.00 %	23.03 %
	Revocação	0.00 %	100.00 %	60.61 %
Fisherfaces	Precisão	0.00 %	13.22 %	6.30 %
	Revocação	0.00 %	100.00 %	41.65 %
SIFT	Precisão	0.00 %	8.56 %	2.09 %
	Revocação	0.00 %	100.00 %	19.48 %
SURF	Precisão	0.00 %	5.88 %	0.95 %
	Revocação	0.00 %	100.00 %	18.63 %

Fonte: Kelvin Salton do Prado, 2017.

Como é possível visualizar na Tabela 10, o método que obteve a maior média de precisão (23.03%) e revocação (60.61%) foi o LBPH, enquanto que o método SURF teve a menor média de precisão (0.95%) e revocação (18.63%). Por outro lado, na Tabela 11 (apresentada a seguir) é possível notar que o método Eigenfaces teve a maior média de precisão (12.77%) e revocação (57.82%), porém as mesmas ainda foram mais baixas que os resultados obtidos pelo método LBPH no experimento anterior (referente à Tabela 10).

Tabela 11 – Resumo da precisão e revocação dos experimentos envolvendo a base de treinamento 6 e o vídeo 11

Método	Precisão e Revocação	Mín.	Máx.	Média
Eigenfaces	Precisão	0.00 %	24.39 %	12.77 %
	Revocação	0.00 %	100.00 %	57.82 %
LBPH	Precisão	0.00 %	33.33 %	7.64 %
	Revocação	0.00 %	100.00 %	53.97 %
Fisherfaces	Precisão	0.00 %	19.05 %	7.90 %
	Revocação	0.00 %	100.00 %	50.96 %
SIFT	Precisão	4.69 %	16.67 %	10.46 %
	Revocação	11.11 %	100.00 %	39.65 %
SURF	Precisão	0.00 %	7.69 %	1.65 %
	Revocação	0.00 %	100.00 %	20.50 %

Fonte: Kelvin Salton do Prado, 2017.

Cabe esclarecer que, apesar de o método LBPH apresentar precisão máxima de 100%, na Tabela 10, isso não significa que o resultado foi de forma geral satisfatório, pois isto ocorre apenas pelo fato de não existirem falsos positivos, ou seja, independente da quantidade de verdadeiros positivos. Neste contexto, mesmo que o método tenha reconhecido corretamente apenas 3 sujeitos (verdadeiros positivos) e nenhum falso positivo (o que ocorreu no experimento utilizando o limiar 98 para o LBPH com a base de treinamento 5 o vídeo de teste 5), a precisão será de 100%, porém o resultado, neste caso, não é considerado expressivo, pois o algoritmo reconheceu corretamente (abaixo do limiar) apenas 3 sujeitos. Ainda, a quantidade de verdadeiros positivos em cada experimento pode ser visualizada pela linha azul encontrada nos gráficos superiores das figuras apresentadas anteriormente nesta seção (e.g. Figuras 44 e 45).

Apesar de os resultados apresentados nas Tabelas 10 e 11 serem interessantes para a finalidade de comparação, não é possível considerar eles como resultados absolutos, pois os valores mínimos, máximos e médios relacionados a precisão e revocação podem ser alterados de acordo com a escolha e a quantidade de limiares utilizados nos experimentos.

Ainda nesta seção, utilizando a análise da curva ROC, é possível comparar a taxa de verdadeiros positivos utilizando diferentes bases de treinamento. A Figura 49 apresenta a curva ROC de 2 experimentos utilizando as bases de treinamento 1 e 6, o vídeo de teste 11 e o método LBPH. Nela, o gráfico da esquerda representa o experimento utilizando a base de treinamento 1 e o gráfico da direita corresponde ao experimento utilizando a base de treinamento 6.

Figura 49 – Curva ROC da comparação utilizando diferentes bases de treinamento.

Fonte: Kelvin Salton do Prado, 2017.

Como é possível visualizar na Figura 49, a curva ROC do experimento utilizando a primeira base de treinamento (gráfico da esquerda) está bem próxima da curva pontilhada vermelha, a qual representa um resultado aleatório, e, em alguns casos, tem um desempenho abaixo de um método aleatório. Por outro lado, é possível notar que a curva ROC apresentada no gráfico da direita, utilizando a sexta base de treinamento, possui um desempenho um pouco melhor se comparado ao resultado apresentado no gráfico da esquerda, sendo que é possível notar que em nem um momento a curva está abaixo de um resultado aleatório. Ainda que não seja possível concluir que a utilização da sexta base de treinamento realmente possa gerar resultados melhores que a primeira base, pois os resultados obtidos neste experimento foram muito próximos e não apresentaram uma grande melhoria, acredita-se que a utilização de múltiplas imagens de treinamento para cada sujeito (o que ocorre na sexta base de treinamento) pode influenciar, geralmente de forma positiva, os resultados obtidos em relação a uma base de treinamento com apenas uma imagem de cada participante (o que ocorre na primeira base).

4.5.4 Resultados do comitê de máquinas

Nesta seção será apresentada uma comparação entre os dois métodos de votação, majoritário e ponderado, utilizados pelo comitê de máquinas desenvolvido neste trabalho.

Como já explicado nas Seções 3.3, 3.3.1 e 3.3.2, dado um conjunto de resultados, os quais foram classificados pelos componentes do comitê de máquinas (neste caso os algoritmos de reconhecimento facial), o voto majoritário basicamente classifica o sujeito de acordo com a maioria dos votos, enquanto o voto ponderado utiliza um sistema de pesos, definido pelo usuário, para classificar o sujeito que tiver maior peso após a somatória dos votos.

Com a intenção de obter resultados mais representativos para realizar a comparação dos métodos de votação, bem como facilitar o entendimento dos resultados, os experimentos apresentados a seguir foram realizados utilizando a sexta base de treinamento e o décimo primeiro vídeo de teste. As bases de treinamento e teste foram selecionadas para realizar os seguintes experimentos pois, como pode ser visto nos resultados apresentados na Seção 4.5.2, geraram resultados mais expressivos e, por este motivo, facilitam a visualização e comparação dos resultados.

Assim como na Seção 4.5.2, nos experimentos apresentados a seguir não são utilizados limiares, ou seja, são medidas apenas as quantidades de resultados corretos e incorretos dos algoritmos. Além disso, assim como nos demais experimentos, foram utilizadas apenas as imagens faciais detectadas, não levando em consideração as imagens que não correspondem a faces ou imagens de sujeitos que não se encontram na base de treinamento. Neste contexto, é possível considerar os resultados corretos como verdadeiros positivos e os incorretos como falsos positivos e utilizar a Equação 30 apresentada na Seção 3.7 para calcular a precisão.

Na Tabela 12, são apresentados os resultados do comitê de máquinas utilizando 3 componentes e voto majoritário. Os componentes foram variados de forma a combinar todos os algoritmos entre si, sem repetições, gerando ao final 10 resultados, os quais podem ser visualizados na coluna denominada “Precisão”. Como é possível visualizar na Tabela 12 apresentada a seguir, a precisão máxima obtida foi de 19.27% utilizando os conjuntos de componentes Eigenfaces, LBPH e SURF, e Eigenfaces, Fisherfaces e SURF. Ainda, é possível verificar que a precisão mínima obtida foi de 9.90% utilizando o conjunto de componentes Fisherfaces, SIFT e SURF.

Tabela 12 – Resultados do comitê de máquinas utilizando 3 componentes e voto majoritário

Componentes do Comitê de Máquinas			Precisão
Eigenfaces	LBPH	Fisherfaces	18.23 %
Eigenfaces	LBPH	SIFT	18.75 %
Eigenfaces	LBPH	SURF	19.27 %
Eigenfaces	Fisherfaces	SIFT	18.75 %
Eigenfaces	Fisherfaces	SURF	19.27 %
Eigenfaces	SIFT	SURF	18.23 %
LBPH	Fisherfaces	SIFT	11.98 %
LBPH	Fisherfaces	SURF	12.50 %
LBPH	SIFT	SURF	10.42 %
Fisherfaces	SIFT	SURF	09.90 %

Fonte: Kelvin Salton do Prado, 2017.

Na Tabela 13 são apresentados os resultados do comitê de máquinas utilizando todos os 5 algoritmos de reconhecimento facial como componentes, e o voto ponderado como sistema de votação. Os valores dos pesos foram variados de 0 a 40 para cada algoritmo independente, porém, cabe lembrar que, no caso em que todos os pesos são iguais, o algoritmo se comporta como o voto majoritário. Como é possível visualizar na Tabela 13 apresentada a seguir, a precisão máxima obtida com os experimentos foi de 18.75%, enquanto a precisão mínima obtida foi de 4.69%. Cabe ainda esclarecer que, neste caso, como estamos variando o peso de apenas 1 componente por vez e os demais componentes possuem pesos igual a 10, deve-se definir valores múltiplos de 10 para o componente que está sendo variado o peso, pois valores diferentes destes não trarão resultados distintos, ou seja, utilizando como exemplo a terceira linha da Tabela 13, a variação do peso do Eigenfaces entre 11 e 19 trará o mesmo resultado que a utilização do peso 20, lembrando que, no sistema de voto ponderado, em caso de empate ganha o resultado que tiver maior quantidade de votos.

Tabela 13 – Resultados do comitê de máquinas utilizando 5 componentes e voto ponderado

Eigenfaces	Pesos				Precisão
	LBP	Fisherfaces	SIFT	SURF	
0	10	10	10	10	06.77 %
10	10	10	10	10	17.19 %
20	10	10	10	10	17.19 %
30	10	10	10	10	18.75 %
40	10	10	10	10	18.75 %
10	0	10	10	10	17.71 %
10	10	10	10	10	17.19 %
10	20	10	10	10	06.25 %
10	30	10	10	10	12.50 %
10	40	10	10	10	11.46 %
10	10	0	10	10	17.71 %
10	10	10	10	10	17.19 %
10	10	20	10	10	11.98 %
10	10	30	10	10	11.46 %
10	10	40	10	10	11.46 %
10	10	10	0	10	17.19 %
10	10	10	10	10	17.19 %
10	10	10	20	10	07.29 %
10	10	10	30	10	05.21 %
10	10	10	40	10	04.69 %
10	10	10	10	0	17.71 %
10	10	10	10	10	17.19 %
10	10	10	10	20	06.77 %
10	10	10	10	30	04.69 %
10	10	10	10	40	05.73 %

Fonte: Kelvin Salton do Prado, 2017.

Comparando os resultados apresentados nas Tabelas 12 e 13 é possível notar que o comitê de máquinas utilizando apenas 3 componentes e o voto majoritário teve melhores resultados em comparação com o comitê de máquinas utilizando 5 componentes e voto ponderado, sendo que no primeiro foi possível atingir uma precisão de 19.27% enquanto no segundo a precisão máxima obtida foi de 18.75%.

Comparando os resultados obtidos com o comitê de máquinas em relação aos resultados apresentados na Tabela 9 da Seção 4.5.2, é possível ver uma pequena melhora nos resultados em relação aos algoritmos individuais, sendo que individualmente, utilizando a base de treinamento 6 e o vídeo 11, o algoritmo Eigenfaces obteve a precisão de 18.75% enquanto com o comitê de máquinas foi possível obter uma precisão de 19.27%.

Com isso, é possível concluir que com o comitê de máquinas foi possível obter uma pequena melhora na precisão, porém não tão significativa. Acredita-se que este resultado

ocorreu pois os componentes do comitê de máquinas precisam ter um bom desempenho individualmente para que o comitê de máquinas obtenha um desempenho melhor, assim como explicado por Lima (2004). Neste caso, apesar da pequena melhora obtida, deve-se levar em consideração que com o comitê de máquinas tem-se também um aumento do custo computacional, pois é necessária a utilização de mais de um componente para compor o comitê, o que, em alguns casos, pode tornar inviável a sua utilização em um sistema de reconhecimento facial em ambiente real.

5 Considerações finais

De acordo com os experimentos realizados, é possível concluir que as taxas de reconhecimento facial foram relativamente baixas, e que isto ocorre, principalmente, devido à baixa qualidade das imagens de teste em comparação com as imagens de treinamento.

Sabe-se que é possível obter boas taxas de reconhecimento facial em ambientes controlados (e.g. (MAO; LI; YIN, 2014; DINIZ et al., 2013)), geralmente onde as imagens de treinamento e de testes possuem boa nitidez, e em lugares onde é possível controlar, mesmo que parcialmente, a orientação da pose e iluminação. Como pode-se visualizar nos trabalhos relacionados (Seção 2.2), é possível obter taxas de reconhecimento acima de 90% (e.g. (DINIZ et al., 2013)), chegando a atingir taxas de até 100% em alguns casos (e.g. (MAO; LI; YIN, 2014)), ou seja, reconhecer corretamente todas as imagens faciais.

Por outro lado, o reconhecimento facial utilizando câmeras de vigilância ainda é um desafio para esta área de pesquisa, sendo que em geral as taxas de reconhecimento são baixas nestes ambientes, assim como pode ser visto no quarto trabalho relacionado (XU; LIU; LI, 2014) apresentado na Seção 2.2, pois geralmente as imagens obtidas são de baixa qualidade e, em muitos casos, se diferem bastante das imagens utilizadas no treinamento, assim como ocorre com os dados utilizados neste trabalho.

Ainda, como visto no trabalho relacionado, Xu, Liu e Li (2014) colocam que a distância da câmera e a resolução das imagens levam a significativas diferenças nas taxas de reconhecimento facial, sendo que, em uma distância de 5 metros foi possível obter uma taxa de reconhecimento de 70% quando utilizadas imagens de uma câmera de alta definição, enquanto na mesma distância, utilizando imagens de uma câmera de vigilância, foi possível obter uma taxa de apenas 11%. Neste trabalho, como o processo de reconhecimento facial é aplicado em diferentes salas, as quais possuem diferentes tamanhos, não é possível calcular exatamente a distância entre os sujeitos e as câmeras, entretanto estima-se que a mesma varia entre 2 e 8 metros. Com base nas distâncias aproximadas é possível concluir que os resultados obtidos neste trabalho foram, de certa forma, semelhantes aos obtidos por Xu, Liu e Li (2014), utilizando uma câmera de vigilância em uma distância de 5 metros, pois foi possível obter uma média de precisão de aproximadamente 8%, como visto nos experimentos apresentados na Seção 4.5.2, sendo que, em alguns vídeos, esta média foi mais baixa e em outros mais alta.

Ademais, a distância entre as câmeras e os alunos pode ser considerada um fator de limitação deste trabalho. Tal distância, somada à baixa qualidade das imagens obtidas por meio das câmeras de segurança, prejudica significativamente o processo de detecção e reconhecimento facial. Notou-se em alguns vídeos que os sujeitos mais ao fundo da sala raramente tem a face detectada pelo algoritmo Viola-Jones e, ainda, quando a mesma é detectada, a imagem facial é de baixa qualidade e, por este motivo, dificilmente o sujeito é reconhecido. Um exemplo deste problema foi apresentado na Tabela 3 da Seção 4.3.3, na qual é possível visualizar que, no vídeo 2, aparecem um total de 8 participantes, porém, apenas 4 deles (os que aparecem mais a frente no vídeo) tem a face detectada e somente 3 tem a face reconhecida pelos algoritmos, de acordo com conferência manual. Cabe lembrar que, como mencionado na Seção 4.3.3, o vídeo 2 foi gravado de um dispositivo móvel durante uma apresentação dos participantes, ou seja, não corresponde a um vídeo obtido das câmeras de segurança da academia. Ainda que a qualidade das imagens obtidas do vídeo 2 pareçam ser moderadamente melhores que as imagens provenientes dos demais vídeos (obtidos de câmeras de segurança), não foi possível notar uma melhora nos resultados, como visto na Seção 4.5.2. Ainda assim, acredita-se que a utilização de câmeras de segurança de alta resolução, juntamente com um conjunto de treinamento que represente bem as imagens de teste, poderiam melhorar significativamente os resultados obtidos no processo de reconhecimento facial.

Além disso, de acordo com os trabalhos apresentados na Seção 2.2, é possível notar que poucos tratam de ambientes reais e semicontrolados, como o exemplo de Xu, Liu e Li (2014), que utilizam imagens de uma câmera de vigilância. Além disso, de acordo com os mesmos autores, a maior parte dos trabalhos envolvendo imagens de baixa resolução baseia-se no que é chamado de *down-sampled low-resolution*, ou seja, imagens originalmente obtidas em alta resolução, porém submetidas a métodos de reescalonamento ou suavização, as quais não são boas substitutas para imagens diretamente capturadas em baixa resolução. Ainda, de acordo com os trabalhos relacionados, mais precisamente com base no Quadro 1 da Seção 2.3, é possível visualizar que os trabalhos utilizam diversas bases de dados (e.g. Honda/UCSD, ScFace, FRGC), as quais foram geralmente geradas em ambientes controlados e com uma boa qualidade de imagens. Nesta mesma seção, é possível notar também que os trabalhos relacionados utilizam diferentes técnicas, não necessariamente voltadas para o reconhecimento facial. Dentre elas, são utilizadas técnicas de agrupamento, de classificação e regressão (e.g. *K-Means*, *Random Forest*, *K-NN*, *K-Star*). Em contraste,

o presente trabalho utiliza uma base de dados própria, criada em um ambiente real e semicontrolado, onde são encontrados diversos problemas (e.g. variação de iluminação e baixa nitidez), o que, por si só, já torna o processo de reconhecimento facial um grande desafio.

Cabe ainda lembrar que foram removidos dos experimentos os dados considerados “ruídos”, ou seja, imagens detectadas que não correspondiam à imagens faciais e imagens faciais que não correspondiam aos participantes desta pesquisa, assim como descrito com maiores detalhes na Seção 4.5.2. Optou-se por remover estes dados com o objetivo de obter resultados mais expressivos, ou seja, que representassem melhor os resultados dos algoritmos de reconhecimento, sem a interferência de falhas oriundas dos algoritmos de detecção de face, além de tornar a base de dados mais semelhante às bases já conhecidas (e.g. Yale Face Database¹), as quais fazem uso apenas de imagens faciais. Dessa forma, é possível dizer que os resultados obtidos são otimistas, no sentido de que pressupõem uma base sem “ruídos”, ou seja, uma precisão de 100% na etapa de detecção facial, bem como a ausência de faces externas à base de treinamento utilizada no processo de reconhecimento facial.

Com isso, de acordo com os resultados encontrados foi possível considerar que as diversas variáveis citadas anteriormente, tais como variação de iluminação, variação de pose, oclusão e baixa nitidez, prejudicam significativamente o desempenho do reconhecimento facial. Além disso, a quantidade e qualidade das imagens faciais utilizadas na etapa de treinamento dos algoritmos estão diretamente ligadas aos resultados obtidos, sendo que quanto mais próximas das imagens de teste forem as imagens de treinamento, melhores são os resultados obtidos.

Ainda, foi possível notar que os parâmetros, tanto aqueles utilizados na detecção facial quanto aqueles utilizados no reconhecimento facial, estão diretamente ligados aos resultados obtidos e que, devido a esta importância, merecem um estudo mais aprofundado e específico, o qual não foi possível desenvolver neste trabalho devido principalmente ao escopo e o prazo.

Finalizando, cabe colocar que os algoritmos utilizados neste trabalho, desenvolvidos em Python e Matlab, estão disponíveis publicamente em um repositório no site Github², e serão mantidos atualizados com correções de *bugs* e melhorias, de forma que podem ser

¹ Link: <http://vision.ucsd.edu/content/yale-face-database>.

² Link para o repositório: <https://github.com/kelvins/Reconhecimento-Facial>

úteis para o desenvolvimento de futuros trabalhos na área de reconhecimento facial, bem como a replicação dos experimentos realizados neste.

6 Conclusões

Primeiramente, cabe aqui recordar brevemente os objetivos deste trabalho, os quais são divididos em objetivo geral e específico.

- O objetivo geral deste trabalho consistiu na comparação de diversas técnicas de reconhecimento facial em um ambiente real, semicontrolado e não colaborativo, com a intenção de identificar a técnica que melhor se adaptasse ao ambiente estudado.
- O objetivo específico, por sua vez, consistiu na análise da possibilidade de realizar o reconhecimento automático de pessoas no ambiente real da academia, utilizando uma ou mais das técnicas de reconhecimento facial estudadas.

Segundo os experimentos realizados e com resultados apresentados na Seção 4.5, é possível concluir que o método LBPH teve o melhor desempenho, ao utilizar os parâmetros padrões da biblioteca OpenCV no ambiente proposto por este trabalho. Por outro lado, é possível notar também que o método Eigenfaces teve um desempenho inferior em diversos experimentos, se comparado aos demais métodos, o que acaba, de certa forma, tornando-se uma opção inviável para a realização do reconhecimento facial no ambiente proposto. Além disso, foi possível notar também que apesar dos métodos SIFT e SURF lidarem relativamente bem com variação de escala e oclusão parcial, eles não apresentaram resultados significativamente melhores que os demais métodos nos experimentos realizados.

Ainda, de acordo com os resultados encontrados, acredita-se não ser possível realizar a detecção automática de forma confiável apenas utilizando os métodos de reconhecimento facial no ambiente atual da academia. Com os experimentos, visualizou-se que os principais fatores que levam a esta conclusão são a baixa qualidade das imagens das câmeras, pois tratam-se de câmeras de vigilância, bem como sua posição que, em geral, não se apresenta de forma favorável para a realização da detecção, e conseqüentemente do reconhecimento facial. Sendo assim, supõe-se que com a utilização de câmeras de alta resolução posicionadas em locais estratégicos, com o objetivo de detectar a face dos sujeitos, poderiam-se obter resultados melhores, e possivelmente utilizar o reconhecimento facial nesse ambiente.

Conclui-se também que, ainda que os resultados encontrados com os experimentos não tenham sido suficientemente bons para a realização da detecção automática dos alunos da academia, o trabalho, de forma geral, alcançou os objetivos propostos inicialmente,

levantou questões e trouxe informações relevantes que podem contribuir para o estado da arte na área de reconhecimento facial.

6.1 *Trabalhos futuros*

Como sugestões de trabalhos futuros, é possível citar:

- Tendo a consciência de que as bases de treinamento utilizadas neste trabalho podem ser uma lacuna que pode ter prejudicado os resultados obtidos, sugere-se, como trabalho futuro, criar uma nova base de treinamento utilizando mais imagens faciais para cada participante, de preferência com variação de posição da face, de resolução e de iluminação, gerando assim uma base de treinamento mais robusta, e que pode representar melhor as imagens faciais obtidas nos vídeos. Se porventura não for possível obter imagens faciais com variações de posição, sugere-se então rotacionar as imagens com a intenção de obter mais imagens que possam representar melhor as imagens obtidas nos vídeos.
- Buscando melhorar a qualidade das imagens obtidas dos vídeos de teste, sugere-se a utilização de técnicas de redimensionamento mais robustas, que permitam aumentar a dimensão da imagem melhorando a qualidade da mesma. Pode-se citar como exemplo a técnica *Rapid and Accurate Image Super Resolution* (RAISR) que, de acordo com os autores, é capaz de produzir uma imagem de maior dimensão produzindo resultados comparáveis ou até melhores que outras técnicas do estado da arte (ROMANO; ISIDORO; MILANFAR, 2017).
- Sugere-se também a análise detalhada dos parâmetros de cada método de reconhecimento facial. Como mencionado na Seção 4.5.1, a biblioteca OpenCV utiliza atualmente parâmetros padrões definidos com base nos artigos originais de cada método, bem como estudos relacionados a estes, porém, acredita-se que a utilização de diferentes parâmetros poderia resultar em diferentes resultados.
- Com a intenção de tornar possíveis e replicáveis os experimentos realizados no trabalho, sugere-se disponibilizar publicamente os dados de treinamento e teste obtidos, de acordo com autorização prévia dos participantes, ou até mesmo utilizar bases de dados públicas, de preferência criadas em ambientes reais, com a finalidade de comparação dos algoritmos e técnicas de reconhecimento facial.

- Outro ponto que pode ser sugerido como trabalho futuro, porém que foge do escopo do trabalho atual, seria reposicionar as câmeras de vigilância da academia alvo do estudo, posicionando-as em um local estratégico de forma a capturar imagens faciais frontais dos indivíduos, buscando assim obter imagens mais adequadas para a realização do reconhecimento facial.

Referências¹

- ADINI, Y.; MOSES, Y.; ULLMAN, S. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 19, n. 7, p. 721–732, 1997. Citado na página 25.
- AHONEN, T.; HADID, A.; PIETIKÄINEN, M. Face recognition with local binary patterns. In: *Computer vision-ECCV 2004*. Prague, Czech Republic: Springer, 2004. p. 469–481. Citado 5 vezes nas páginas 28, 45, 46, 47 e 104.
- AHONEN, T.; HADID, A.; PIETIKAINEN, M. Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 28, n. 12, p. 2037–2041, 2006. Citado 2 vezes nas páginas 47 e 48.
- AMARAL, V. do; GIRALDI, G. A.; THOMAZ, C. E. Lbp estatístico aplicado ao reconhecimento de expressões faciais. 2013. Citado 2 vezes nas páginas 28 e 45.
- ARANDJELOVIĆ, O.; CIPOLLA, R. Face recognition from video using the generic shape-illumination manifold. *Computer Vision-ECCV 2006*, Springer, p. 27–40, 2006. Citado na página 30.
- ARANDJELOVIC, O.; CIPOLLA, R. Colour invariants for machine face recognition. In: IEEE. *8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008*. Amsterdam, Netherlands, 2008. p. 1–8. Citado na página 30.
- ARANDJELOVIĆ, O.; CIPOLLA, R. Achieving robust face recognition from video by combining a weak photometric model and a learnt generic face invariant. *Pattern Recognition*, Elsevier, v. 46, n. 1, p. 9–23, 2013. Citado na página 30.
- ARANDJELOVIĆ, O.; HAMMOUD, R.; CIPOLLA, R. Thermal and reflectance based personal identification methodology under variable illumination. *Pattern Recognition*, Elsevier, v. 43, n. 5, p. 1801–1813, 2010. Citado na página 30.
- ARAUJO, G. M. *Algoritmo para reconhecimento de características faciais baseado em filtros de correlação*. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, 2010. Citado 2 vezes nas páginas 27 e 39.
- BAK, S. et al. Boosted human re-identification using riemannian manifolds. *Image and Vision Computing*, Elsevier, v. 30, n. 6, p. 443–452, 2012. Citado na página 25.
- BAY, H. et al. Speeded-up robust features (surf). *Computer vision and image understanding*, Elsevier, v. 110, n. 3, p. 346–359, 2008. Citado 4 vezes nas páginas 59, 60, 61 e 62.
- BAY, H.; TUYTELAARS, T.; GOOL, L. V. Surf: Speeded up robust features. In: SPRINGER. *European conference on computer vision*. Graz, Austria, 2006. p. 404–417. Citado 3 vezes nas páginas 28, 58 e 59.

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

BELHUMEUR, P. N.; HESPANHA, J. P.; KRIEGMAN, D. J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 19, n. 7, p. 711–720, 1997. Citado 2 vezes nas páginas 49 e 51.

BELO, F. A. W. Desenvolvimento de algoritmos de exploração e mapeamento visual para robôs móveis de baixo custo. *Rio de Janeiro*, 2006. Citado na página 53.

BISWAS, S. et al. Pose-robust recognition of low-resolution face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 35, n. 12, p. 3037–3049, 2013. Citado 4 vezes nas páginas 20, 27, 32 e 35.

BRADSKI, G.; KAEHLER, A. *Learning OpenCV: Computer vision with the OpenCV library*. [S.l.]: "O'Reilly Media, Inc.", 2008. Citado na página 68.

BUCKLAND, M.; GEY, F. The relationship between recall and precision. *Journal of the American society for information science*, American Documentation Institute, v. 45, n. 1, p. 12, 1994. Citado na página 75.

BURRUS, C. S.; GOPINATH, R. A.; GUO, H. Introduction to wavelets and wavelet transforms: a primer. Prentice-Hall, Inc., 1997. Citado na página 69.

BURRUS, C. S.; GOPINATH, R. A.; GUO, H. *Introduction to wavelets and wavelet transforms: a primer*. Houston, Texas: Upper Saddle River: Prentice Hall, 1998. Citado 2 vezes nas páginas 69 e 70.

CHAO, W.-L. Face recognition. *GICE, National Taiwan University*, 2007. Citado na página 72.

COSTA, D. M. M. d. *Ensemble baseado em métodos de Kernel para reconhecimento biométrico multimodal*. Dissertação (Mestrado) — Universidade de São Paulo, 2016. Citado 5 vezes nas páginas 39, 40, 69, 70 e 85.

DHARAVATH, K.; TALUKDAR, F.; LASKAR, R. Study on biometric authentication systems, challenges and future trends: A review. In: IEEE. *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*. Tamilnadu, India, 2013. p. 1–7. Citado na página 40.

DINIZ, F. A. et al. Redface: um sistema de reconhecimento facial baseado em técnicas de análise de componentes principais e autofaces. *Revista Brasileira de Computação Aplicada*, v. 5, n. 1, p. 42–54, 2013. Citado 6 vezes nas páginas 20, 27, 28, 33, 36 e 133.

DOURADO, W. B. Avaliação de técnicas de interpolação de imagens digitais. Universidade Estadual Paulista (UNESP), 2014. Citado na página 77.

FAWCETT, T. An introduction to roc analysis. *Pattern recognition letters*, Elsevier, v. 27, n. 8, p. 861–874, 2006. Citado 3 vezes nas páginas 74, 75 e 76.

FIGUEREDO, M. *Reconhecimento de faces aplicado ao problema de pessoas desaparecidas - Estudo de caso do Eigenface*. Dissertação (Mestrado) — SENAI CIMATEC, Salvador, 2011. Citado na página 49.

GONZÁLES, G. L. G. *Aplicação da Técnica SIFT para Determinação de Campos de Deformações de Materiais*. Dissertação (Mestrado) — PUC-Rio, 2010. Citado 4 vezes nas páginas 52, 53, 54 e 56.

GUTTA, S.; WECHSLER, H. Face recognition using hybrid classifier systems. In: IEEE. *IEEE International Conference on Neural Networks*. Washington, D.C., USA, 1996. v. 2, p. 1017–1022. Citado na página 65.

HAN, D. Comparison of commonly used image interpolation methods. *ICCSEE, Hangzhou, China*, p. 1556–1559, 2013. Citado na página 77.

HANSEN, L. K.; SALAMON, P. Neural network ensembles. *IEEE transactions on pattern analysis and machine intelligence*, v. 12, p. 993–1001, 1990. Citado 2 vezes nas páginas 64 e 117.

HEO, J.; SAVVIDES, M.; VIJAYAKUMAR, B. Performance evaluation of face recognition using visual and thermal imagery with advanced correlation filters. In: IEEE. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops*. San Diego, CA, USA, 2005. p. 9–9. Citado na página 30.

HUANG, F. J. et al. Pose invariant face recognition. In: IEEE. *Proceedings. Fourth IEEE International Conference on Automatic Face and Gesture Recognition*. Grenoble, France, 2000. p. 245–250. Citado na página 65.

JAIN, A. K.; LI, S. Z. *Handbook of face recognition*. [S.l.]: Springer, 2005. v. 1. Citado na página 19.

KAMI, G. J. d. C. Análise de técnicas de reconhecimento de padrões para a identificação biométrica de usuários em aplicações web utilizando faces a partir de vídeos. Universidade Estadual Paulista (UNESP), 2011. Citado na página 50.

LIMA, C. A. de M. *Comitê de Máquinas: uma abordagem unificada empregando máquinas de vetores-suporte*. Tese (Doutorado) — Universidade Estadual de Campinas, 2004. Citado 4 vezes nas páginas 64, 116, 117 e 132.

LOUIS, W.; PLATANIOTIS, K. N. Frontal face detection for surveillance purposes using dual local binary patterns features. In: IEEE. *IEEE International Conference on Image Processing (ICIP), 2010 17th*. Hong Kong, 2010. p. 3809–3812. Citado na página 48.

LOUIS, W.; PLATANIOTIS, K. N.; RO, Y. M. Enhanced weakly trained frontal face detector for surveillance purposes. In: IEEE. *IEEE International Conference on Fuzzy Systems (FUZZ)*. Barcelona, Spain, 2010. p. 1–8. Citado na página 48.

LOWE, D. G. Object recognition from local scale-invariant features. In: IEEE. *The proceedings of the seventh IEEE international conference on Computer vision*. Kerkyra, Greece, 1999. v. 2, p. 1150–1157. Citado 2 vezes nas páginas 28 e 52.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, Springer, v. 60, n. 2, p. 91–110, 2004. Citado 7 vezes nas páginas 52, 53, 54, 55, 56, 57 e 58.

MACHADO, B. B. et al. Implementação de um algoritmo de reconhecimento facial usando eigenface. *e-xacta*, v. 2, n. 2, 2009. Citado na página 28.

MACHADO, W. et al. Estudo e proposta de adaptação do algoritmo sift em relação ao problema de iluminação em imagens. In: *IV Workshop de Visão Computacional. Anais do WVC*. [S.l.: s.n.], 2008. Citado na página 28.

MALLAT, S. *A wavelet tour of signal processing: the sparse way*. [S.l.]: Academic press, 2008. Citado na página 69.

MAO, Y.; LI, H.; YIN, Z. Who missed the class?—unifying multi-face detection, tracking and recognition in videos. In: IEEE. *IEEE International Conference on Multimedia and Expo - ICME*. Chengdu, China, 2014. p. 1–6. Citado 5 vezes nas páginas 20, 27, 30, 35 e 133.

MARTINEZ, E. Z.; LOUZADA-NETO, F.; PEREIRA, d. B. B. A curva roc para testes diagnósticos. *Cad Saúde Coletiva*, v. 11, n. 1, p. 7–31, 2003. Citado na página 71.

MIAN, P. et al. A systematic review process for software engineering. In: *ESELAW'05: 2nd Experimental Software Engineering Latin American Workshop*. [S.l.: s.n.], 2005. Citado na página 29.

MULCAHY, C. Image compression using the haar wavelet transform. *Spelman Science and Mathematics Journal*, v. 1, n. 1, p. 22–31, 1997. Citado na página 85.

OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, Elsevier, v. 29, n. 1, p. 51–59, 1996. Citado na página 45.

OJALA, T.; PIETIKAINEN, M.; MAENPAA, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, IEEE, v. 24, n. 7, p. 971–987, 2002. Citado na página 48.

OPENCV. *OpenCV Documentation*. [S.l.], 2000. [Acesso em 18 de maio de 2016]. Disponível em: <<http://docs.opencv.org/2.4/index.html>>. Citado 10 vezes nas páginas 28, 62, 68, 90, 91, 100, 103, 104, 105 e 106.

PEDERSEN, J. T. Study group surf: Feature detection & description. *Department of Computer Science, Aarhus University*, 2011. Citado 4 vezes nas páginas 58, 60, 61 e 62.

PENTLAND, A. et al. View-based and modular eigenspaces for face recognition. In: *Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: [s.n.], 1994. v. 94, p. 84–91. Citado na página 74.

POULARIKAS, A. D. *Transforms and applications handbook*. [S.l.]: CRC press, 2010. Citado na página 70.

PRADO, K. S. et al. Automatic facial recognition: A systematic review on the problem of light variation. In: SPRINGER. *International Conference on Human-Computer Interaction*. Toronto, Canada, 2016. p. 211–221. Citado 2 vezes nas páginas 30 e 38.

PRASHANTH, H.; SHASHIDHARA, H.; KN, B. M. Image scaling comparison using universal image quality index. In: IEEE. *International Conference on Advances in Computing, Control and Telecommunication Technologies, 2009. ACT'09*. Trivandrum, Kerala, India, 2009. p. 859–863. Citado 2 vezes nas páginas 77 e 79.

- REISER, B.; FARAGGI, D. Confidence intervals for the generalized roc criterion. *Biometrics*, JSTOR, p. 644–652, 1997. Citado na página 71.
- RIBEIRO, I.; CHIACHIA, G.; MARANA, A. N. Reconhecimento de faces utilizando análise de componentes principais e a transformada census. In: *VI Workshop de Visão Computacional*. Presidente Prudente: FCT/UNESP, 2010. Citado na página 74.
- ROMANO, Y.; ISIDORO, J.; MILANFAR, P. Rair: rapid and accurate image super resolution. *IEEE Transactions on Computational Imaging*, IEEE, v. 3, n. 1, p. 110–125, 2017. Citado na página 138.
- SANTANA, L. M. Q. de; ROCHA, F. G. Processo de detecção facial utilizando viola; jones. *Interfaces Científicas-Exatas e Tecnológicas*, v. 1, n. 1, p. 35–40, 2015. Citado na página 40.
- SANTOS, A. R. *Identificação de faces humanas através de PCA-LDA e redes neurais som*. Dissertação (Mestrado) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2005. Citado na página 49.
- SILVA, A. F. d. et al. Reconhecimento de faces via pca: análise de desempenho. Universidade Federal de Uberlândia, 2006. Citado na página 51.
- SILVA, F. A. da et al. Uma metodologia para detectar e reconhecer placas de sinalização de trânsito. In: *VIII Workshop de Visão Computacional*. [S.l.: s.n.], 2012. Citado 2 vezes nas páginas 55 e 56.
- SOUZA, F. L. d. Classificador fisherface fuzzy para o reconhecimento de faces. Universidade Estadual Paulista (UNESP), 2014. Citado 4 vezes nas páginas 28, 49, 50 e 51.
- SUGANYA, S.; MENAKA, D. Performance evaluation of face recognition algorithms. *International Journal on Recent and Innovation Trends in Computing and Communication*, v. 2, n. 1, 2014. Citado na página 74.
- SZELISKI, R. *Computer vision: algorithms and applications*. [S.l.]: Springer Science & Business Media, 2010. Citado 2 vezes nas páginas 37 e 41.
- TANG, H.-M.; LYU, M. R.; KING, I. Face recognition committee machines: dynamic vs. static structures. In: *12th International Conference on Image Analysis and Processing*. Mantova, Italy: IEEE, 2003. p. 121–126. Citado na página 51.
- TURK, M.; PENTLAND, A. Eigenfaces for recognition. *Journal of cognitive neuroscience*, MIT Press, v. 3, n. 1, p. 71–86, 1991. Citado 5 vezes nas páginas 41, 42, 43, 44 e 45.
- VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: *IEEE. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Kauai, Hawaii, USA, 2001. v. 1, p. I–511. Citado 4 vezes nas páginas 38, 39, 59 e 90.
- VIOLA, P.; JONES, M. J. Robust real-time face detection. *International journal of computer vision*, Springer, v. 57, n. 2, p. 137–154, 2004. Citado na página 40.

WEI, X. et al. Dynamic image-to-class warping for occluded face recognition. *IEEE Transactions on Information Forensics and Security*, IEEE, v. 9, n. 12, p. 2035–2050, 2014. Citado na página 25.

XU, X.; LIU, W.; LI, L. Low resolution face recognition in surveillance systems. *Journal of Computer and Communications*, Scientific Research Publishing, v. 2, n. 02, p. 70, 2014. Citado 6 vezes nas páginas 20, 33, 34, 36, 133 e 134.

YANG, M.-H. *Face recognition using kernel fisherfaces*. United States of America: Google Patents, 2006. US Patent 7,054,468. Citado na página 51.

YANG, M.-H.; KRIEGMAN, D. J.; AHUJA, N. Detecting faces in images: A survey. *IEEE Transactions on pattern analysis and machine intelligence*, IEEE, v. 24, n. 1, p. 34–58, 2002. Citado na página 37.

ZHAO, W. et al. Face recognition: A literature survey. *Acm Computing Surveys (CSUR)*, ACM, v. 35, n. 4, p. 399–458, 2003. Citado na página 19.

ZWEIG, M. H.; CAMPBELL, G. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry*, Am Assoc Clin Chem, v. 39, n. 4, p. 561–577, 1993. Citado na página 71.

Apêndice A – Protocolo da Revisão Sistemática

Protocolo da Revisão Sistemática

Objetivo:

Esta revisão tem como objetivo localizar o estado da arte na área de processamento de imagens, mais precisamente na linha de pesquisa de reconhecimento facial. Tem como foco identificar materiais relevantes à linha de reconhecimento facial em ambientes não controlados e com variação de iluminação.

Estudos Iniciais:

Inicialmente foi realizada uma pequena busca, não formalizada, nas bases de dados da IEEE e ACM, com o objetivo de localizar artigos relacionados ao reconhecimento facial, os quais foram utilizados como base para a análise exploratória. Nesta etapa foi possível localizar os seguintes artigos:

- Face Recognition: The Problem of Compensating for Changes in Illumination Direction.
- Pose-Robust Recognition of Low-Resolution Face Images.
- Low Resolution Face Recognition in Surveillance Systems.
- Who Missed the Class? Unifying Multi-Face Detection, Tracking, and Recognition in Videos.

Estes artigos foram analisados para extração de palavras chaves que foram utilizadas durante as buscas e seleção de fontes.

Formulação da pergunta:

Quais as principais técnicas utilizadas para identificação facial em ambientes não controlados e com variação de iluminação?

Intervenção: Será realizado um levantamento de referencial teórico através de materiais científicos para a compreensão dos métodos e técnicas mais utilizadas para o processo de reconhecimento facial.

Controle: Artigos científicos, teses e dissertações sobre a aplicação de técnicas de reconhecimento facial.

População: Trabalhos relacionados à linha de reconhecimento facial que utilizam câmeras de baixa resolução em ambientes não controlados e com variação de luminosidade.

Resultados: Conhecimento aprofundado das técnicas que envolvem reconhecimento facial em ambientes não controlados, bem como uma base teórica sobre a área de processamento de imagens.

Aplicação: Pesquisadores da área de processamento de imagens e da linha de reconhecimento facial que tenham interesse em aplicar as técnicas em ambientes reais.

Critérios de seleção de fontes:

Métodos de busca de fontes: Utilização das bibliotecas digitais (primeiramente Portal ACM e IEEE Xplore, e futuramente, no caso da continuação deste projeto, Portal de Periódicos CAPES (apenas conteúdo de acesso público) e Banco de Dados Bibliográficos da USP), bem como anais de eventos da área de processamento de imagens.

Palavras-chave identificadas (inglês): image processing, face recognition, face detection, face tracking, multiple object recognition, low resolution, surveillance camera, illumination invariance, pose invariant, person recognition.

Palavras-chave recomendadas (português): processamento de imagens, reconhecimento facial, detecção facial, baixa resolução, câmeras de vigilância, variação de pose, variação de iluminação. Observação: as palavras-chave apresentadas neste parágrafo foram geradas com base nas palavras-chave identificadas em trabalhos de língua inglesa, sendo então recomendadas para a realização de buscas por materiais de língua portuguesa em trabalhos futuros.

Tipos de materiais: Artigos científicos publicados em conferências ou periódicos da área de computação.

Idiomas: Serão primeiramente localizados materiais escritos em língua inglesa, pois uma grande parte dos trabalhos científicos se encontra neste idioma. Se possível, futuramente pretende-se realizar buscas em materiais de língua portuguesa, com o objetivo de abranger a produção científica brasileira relacionada ao tema.

Critérios de inclusão dos artigos:

- a) Trabalhos que definirem técnicas de reconhecimento facial.
- b) Trabalhos que definem conceitos de processamento de imagens e de reconhecimento facial.
- c) Trabalhos que se relacionem com o reconhecimento facial em ambientes reais, não controlados e com variação de iluminação.
- d) Trabalhos que envolvem o reconhecimento facial utilizando câmeras de baixa resolução.

Critérios de exclusão dos artigos:

- a) Trabalhos que não tenham relação com o reconhecimento facial.
- b) Possíveis trabalhos que utilizam técnicas de reconhecimento facial, porém em ambientes controlados como, por exemplo, onde há boa iluminação e utilizam câmeras de alta resolução.
- c) Trabalhos que apresentem avaliação superficial, sem apresentar os métodos e técnicas utilizadas.
- d) Trabalhos que realizam o reconhecimento facial utilizando modelos 3D.
- e) Trabalhos que não tenham sido avaliados por seus pares.
- f) Trabalhos que não envolvam o problema da variação de iluminação.

Critérios de qualidade dos estudos primários:

Como critério de qualidade serão avaliadas as técnicas utilizadas nos trabalhos e sua relevância de acordo com o objetivo da revisão. Também será verificado se o trabalho trata-se de uma aplicação teórica ou prática, sendo possível assim encontrar aplicações reais das técnicas utilizadas.

Processo de seleção dos estudos primários:

Serão construídas '*strings*' de buscas de acordo com as palavras-chave. Estas '*strings*' serão submetidas a motores de busca em bibliotecas online como, por exemplo, Portal ACM e IEEE Xplore. Os títulos e resumos dos trabalhos encontrados serão analisados para verificar sua relevância, sendo os trabalhos aceitos ou descartados de acordo com os critérios de inclusão e exclusão já apresentados. Em caso de dúvidas na inclusão ou exclusão de um artigo, será analisada também a sua conclusão. Ainda em caso de dúvidas, os demais envolvidos direta ou indiretamente com este projeto, tais como orientador, coorientador ou professores do Programa de Pós-Graduação em Sistemas de Informação (PPgSI), poderão ser consultados para avaliarem a inclusão ou exclusão do artigo.

Avaliação da qualidade dos estudos primários:

Serão considerados para inclusão os artigos que atinjam ao menos um dos critérios de inclusão. Caso um artigo atinja um ou mais critérios de inclusão e um de exclusão, será necessário avaliar o artigo com mais precisão e se necessário pedir a avaliação aos demais envolvidos no projeto.

Estratégia de extração de informação:

Primeiramente os artigos encontrados serão analisados de acordo com o título, o resumo, e se necessário, a conclusão, para então serem incluídos ou não na extração de dados. Os artigos aceitos no processo de seleção dos estudos primários deverão ser lidos integralmente, sendo então feito um resumo para cada artigo, destacando não só as técnicas utilizadas, mas também o ambiente envolvido e os resultados obtidos.

Sumarização dos resultados:

Devem ser extraídos de cada trabalho as técnicas utilizadas, os resultados encontrados e as variáveis envolvidas no trabalho. Ao término da revisão, deverá ser realizada uma análise mostrando as principais vantagens das técnicas encontradas. Também deverão ser levantadas possíveis lacunas existentes no estado da arte que poderão ser melhor exploradas.

Apêndice B – Condução da Revisão na Base da ACM

CONDUÇÃO DA REVISÃO SISTEMÁTICA

Fonte: Portal ACM (<http://dl.acm.org/>)

Data de busca: 12/05/2015

Palavras-chave utilizadas: face recognition, face detection, face tracking, low resolution, surveillance camera, illumination invariance, pose invariant, person recognition.

String utilizada: ((*face recognition OR face detection OR face tracking OR person recognition*) AND (*low resolution OR surveillance camera*) AND (*illumination invariance OR pose invariant*))

Total de artigos encontrados: 77

Total de artigos incluídos: 9

Organização dos artigos encontrados: Os artigos encontrados foram catalogados com o auxílio do software *Zotero*¹, e a utilização do padrão bibliográfico “*ACM SIGCHI Proceedings (2016)*”.

Lista de artigos encontrados:

- [1] Abdelkader, M.F., Roy-Chowdhury, A.K., Chellappa, R. and Akdemir, U. 2008. Activity Representation Using 3D Shape Models. *J. Image Video Process.* 2008, (Jan. 2008), 5:1–5:16.
- [2] Aggarwal, J.K. and Ryoo, M.S. 2011. Human Activity Analysis: A Review. *ACM Comput. Surv.* 43, 3 (Apr. 2011), 16:1–16:43.
- [3] Arandjelović, O. and Cipolla, R. 2013. Achieving Robust Face Recognition from Video by Combining a Weak Photometric Model and a Learnt Generic Face Invariant. *Pattern Recogn.* 46, 1 (Jan. 2013), 9–23.
- [4] Arandjelović, O. and Cipolla, R. 2009. A Pose-wise Linear Illumination Manifold Model for Face Recognition Using Video. *Comput. Vis. Image Underst.* 113, 1 (Jan. 2009), 113–125.
- [5] Arandjelović, O. and Cipolla, R. 2006. Face Recognition from Video Using the Generic Shape-illumination Manifold. *Proceedings of the 9th European Conference on Computer Vision - Volume Part IV* (Berlin, Heidelberg, 2006), 27–40.
- [6] Arandjelović, O., Hammoud, R. and Cipolla, R. 2010. Thermal and Reflectance Based Personal Identification Methodology Under Variable Illumination. *Pattern Recogn.* 43, 5 (May

¹ Website: <https://www.zotero.org/>

2010), 1801–1813.

[7] Baskin, S., Charpiat, G., Corvée, E., Brémond, F. and Thonnat, M. 2012. Learning to Match Appearances by Correlations in a Covariance Metric Space. *Proceedings of the 12th European Conference on Computer Vision - Volume Part III* (Berlin, Heidelberg, 2012), 806–820.

[8] Baltieri, D., Vezzani, R. and Cucchiara, R. 2015. Mapping Appearance Descriptors on 3D Body Models for People Re-identification. *Int. J. Comput. Vision.* 111, 3 (Feb. 2015), 345–364.

[9] Bazzani, L., Cristani, M. and Murino, V. 2013. Symmetry-driven Accumulation of Local Features for Human Characterization and Re-identification. *Comput. Vis. Image Underst.* 117, 2 (Feb. 2013), 130–144.

[10] Bedagkar-Gala, A. and Shah, S.K. 2012. Part-based Spatio-temporal Model for Multi-person Re-identification. *Pattern Recogn. Lett.* 33, 14 (Oct. 2012), 1908–1915.

[11] Bicego, M., Grosso, E., Lagorio, A., Brelstaff, G., Brodo, L. and Tistarelli, M. 2008. Distinctiveness of Faces: A Computational Approach. *ACM Trans. Appl. Percept.* 5, 2 (May 2008), 11:1–11:18.

[12] Bk, S., Corvée, E., Brémond, F. and Thonnat, M. 2012. Boosted Human Re-identification Using Riemannian Manifolds. *Image Vision Comput.* 30, 6-7 (Jun. 2012), 443–452.

[13] Bregonzio, M., Xiang, T. and Gong, S. 2012. Fusing Appearance and Distribution Information of Interest Points for Action Recognition. *Pattern Recogn.* 45, 3 (Mar. 2012), 1220–1234.

[14] Cantoni, V., Levialdi, S. and Zavidovique, B. 2011. *3C Vision: Cues, Context and Channels*. Elsevier Science Publishers B. V.

[15] Celik, H., Hanjalic, A. and Hendriks, E.A. 2009. Unsupervised and Simultaneous Training of Multiple Object Detectors from Unlabeled Surveillance Video. *Comput. Vis. Image Underst.* 113, 10 (Oct. 2009), 1076–1094.

[16] Cheddad, A., Mohamad, D. and Manaf, A.A. 2008. Exploiting Voronoi Diagram Properties in Face Segmentation and Feature Extraction. *Pattern Recogn.* 41, 12 (Dec. 2008), 3842–3859.

- [17] Chellappa, R., Sankaranarayanan, A.C., Veeraraghavan, A. and Turaga, P. 2010. Statistical Methods and Models for Video-Based Tracking, Modeling, and Recognition. *Found. Trends Signal Process.* 3, 1–2 (Jan. 2010), 1–151.
- [18] Chen, C., Zhuang, Y. and Xiao, J. 2010. Silhouette Representation and Matching for 3D Pose Discrimination - A Comparative Study. *Image Vision Comput.* 28, 4 (Apr. 2010), 654–667.
- [19] Chen, L., Wei, H. and Ferryman, J. 2013. A Survey of Human Motion Analysis Using Depth Imagery. *Pattern Recogn. Lett.* 34, 15 (Nov. 2013), 1995–2006.
- [20] Connolly, J.-F., Granger, E. and Sabourin, R. 2012. An Adaptive Classification System for Video-based Face Recognition. *Inf. Sci.* 192, (Jun. 2012), 50–70.
- [21] Connolly, J.-F., Granger, E. and Sabourin, R. 2013. Dynamic Multi-objective Evolution of Classifier Ensembles for Video Face Recognition. *Appl. Soft Comput.* 13, 6 (Jun. 2013), 3149–3166.
- [22] Connolly, J.-F., Granger, E. and Sabourin, R. 2012. Evolution of Heterogeneous Ensembles Through Dynamic Particle Swarm Optimization for Video-based Face Recognition. *Pattern Recogn.* 45, 7 (Jul. 2012), 2460–2477.
- [23] Diraco, G., Leone, A. and Siciliano, P. 2010. An Active Vision System for Fall Detection and Posture Recognition in Elderly Healthcare. *Proceedings of the Conference on Design, Automation and Test in Europe* (3001 Leuven, Belgium, Belgium, 2010), 1536–1541.
- [24] Dragon, R., Fenzi, M., Siberski, W., Rosenhahn, B. and Ostermann, J. 2012. Towards Feature-based Situation Assessment for Airport Apron Video Surveillance. *Proceedings of the 15th International Conference on Theoretical Foundations of Computer Vision: Outdoor and Large-scale Real-world Scene Analysis* (Berlin, Heidelberg, 2012), 110–130.
- [25] Eichner, M., Marin-Jimenez, M., Zisserman, A. and Ferrari, V. 2012. 2D Articulated Human Pose Estimation and Retrieval in (Almost) Unconstrained Still Images. *Int. J. Comput. Vision.* 99, 2 (Sep. 2012), 190–214.
- [26] Elguebaly, T. and Bouguila, N. 2013. Simultaneous Bayesian Clustering and Feature Selection Using RJMCMC-based Learning of Finite Generalized Dirichlet Mixture Models. *Signal Process.* 93, 6 (Jun. 2013), 1531–1546.

- [27] Fuentes-Pacheco, J., Ruiz-Ascencio, J. and Rendón-Mancha, J.M. 2015. Visual Simultaneous Localization and Mapping: A Survey. *Artif. Intell. Rev.* 43, 1 (Jan. 2015), 55–81.
- [28] Gade, L., Krishna, S. and Panchanathan, S. 2009. Person Localization Using a Wearable Camera Towards Enhancing Social Interactions for Individuals with Visual Impairment. *Proceedings of the 1st ACM SIGMM International Workshop on Media Studies and Implementations That Help Improving Access to Disabled Users* (New York, NY, USA, 2009), 53–62.
- [29] Girisha, R. and Murali, S. 2011. Tracking Humans Using Novel Optical Flow Algorithm for Surveillance Videos. *Proceedings of the Fourth Annual ACM Bangalore Conference* (New York, NY, USA, 2011), 7:1–7:8.
- [30] Goffredo, M., Bouchrika, I., Carter, J.N. and Nixon, M.S. 2010. Performance Analysis for Automated Gait Extraction and Recognition in Multi-camera Surveillance. *Multimedia Tools Appl.* 50, 1 (Oct. 2010), 75–94.
- [31] Goffredo, M., Bouchrika, I., Carter, J.N. and Nixon, M.S. 2008. Performance Analysis for Gait in Camera Networks. *Proceedings of the 1st ACM Workshop on Analysis and Retrieval of Events/Actions and Workflows in Video Streams* (New York, NY, USA, 2008), 73–80.
- [32] Goudelis, G., Karpouzis, K. and Kollias, S. 2013. Exploring Trace Transform for Robust Human Action Recognition. *Pattern Recogn.* 46, 12 (Dec. 2013), 3238–3248.
- [33] Gouet-Brunet, V. and Lameyre, B. 2008. Object Recognition and Segmentation in Videos by Connecting Heterogeneous Visual Features. *Comput. Vis. Image Underst.* 111, 1 (Jul. 2008), 86–109.
- [34] Grgic, M., Nappi, M. and Wechsler, H. 2012. Editorial: Novel Pattern Recognition-based Methods for Re-identification in Biometric Context. *Pattern Recogn. Lett.* 33, 14 (Oct. 2012), 1817–1819.
- [35] Hwu, W.W. 2011. *GPU Computing Gems Emerald Edition*. Morgan Kaufmann Publishers Inc.
- [36] Islam, S.M.S., Bennamoun, M., Owens, R.A. and Davies, R. 2012. A Review of Recent Advances in 3D Ear- and Expression-invariant Face Biometrics. *ACM Comput. Surv.* 44, 3 (Jun. 2012), 14:1–14:34.

- [37] Jaiswal, S., Bhadauria, S.S., Jadon, R.S. and Divakar, T.K. 2010. Brief Description of Image Based 3D Face Recognition Methods. *3D Res.* 1, 4 (Dec. 2010), 17:1–17:14.
- [38] Ji, X.-F., Wu, Q.-Q., Ju, Z.-J. and Wang, Y.-Y. 2014. Study of Human Action Recognition Based on Improved Spatio-temporal Features. *Int. J. Autom. Comput.* 11, 5 (Oct. 2014), 500–509.
- [39] Kortum, P. 2008. *HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces*. Morgan Kaufmann Publishers Inc.
- [40] Kotsia, I., Guo, W. and Patras, I. 2012. Higher Rank Support Tensor Machines for Visual Recognition. *Pattern Recogn.* 45, 12 (Dec. 2012), 4192–4203.
- [41] Kragic, D. and Vincze, M. 2009. Vision for Robotics. *Found. Trends Robot.* 1, 1 (Jan. 2009), 1–78.
- [42] Layne, R., Hospedales, T.M. and Gong, S. 2012. Towards Person Identification and Re-identification with Attributes. *Proceedings of the 12th International Conference on Computer Vision - Volume Part I* (Berlin, Heidelberg, 2012), 402–412.
- [43] Leo, M., Mazzeo, P.L., Nitti, M. and Spagnolo, P. 2013. Accurate Ball Detection in Soccer Images Using Probabilistic Analysis of Salient Regions. *Mach. Vision Appl.* 24, 8 (Nov. 2013), 1561–1574.
- [44] Liu, C. and Yuen, P.C. 2010. Human Action Recognition Using Boosted EigenActions. *Image Vision Comput.* 28, 5 (May 2010), 825–835.
- [45] Liwicki, S., Tzimiropoulos, G., Zafeiriou, S. and Pantic, M. 2013. Euler Principal Component Analysis. *Int. J. Comput. Vision.* 101, 3 (Feb. 2013), 498–518.
- [46] Lu, W.-L., Okuma, K. and Little, J.J. 2009. Tracking and Recognizing Actions of Multiple Hockey Players Using the Boosted Particle Filter. *Image Vision Comput.* 27, 1-2 (Jan. 2009), 189–205.
- [47] Moeslund, T.B., Hilton, A. and Krüger, V. 2006. A Survey of Advances in Vision-based Human Motion Capture and Analysis. *Comput. Vis. Image Underst.* 104, 2 (Nov. 2006), 90–126.

- [48] Moore, S. and Bowden, R. 2011. Local Binary Patterns for Multi-view Facial Expression Recognition. *Comput. Vis. Image Underst.* 115, 4 (Apr. 2011), 541–558.
- [49] Nasrollahi, K. and Moeslund, T.B. 2014. Super-resolution: A Comprehensive Survey. *Mach. Vision Appl.* 25, 6 (Aug. 2014), 1423–1468.
- [50] Padam Priyal, S. and Bora, P.K. 2013. A Robust Static Hand Gesture Recognition System Using Geometry Based Normalizations and Krawtchouk Moments. *Pattern Recogn.* 46, 8 (Aug. 2013), 2202–2219.
- [51] Pinz, A. 2005. Object Categorization. *Found. Trends. Comput. Graph. Vis.* 1, 4 (Dec. 2005), 255–353.
- [52] Preim, B. and Botha, C.P. 2013. *Visual Computing for Medicine: Theory, Algorithms, and Applications*. Morgan Kaufmann Publishers Inc.
- [53] Razzaghi, P., Palhang, M. and Gheissari, N. 2013. A New Invariant Descriptor for Action Recognition Based on Spherical Harmonics. *Pattern Anal. Appl.* 16, 4 (Nov. 2013), 507–518.
- [54] Ren, W., Singh, S., Singh, M. and Zhu, Y.S. 2009. State-of-the-art on Spatio-temporal Information-based Video Retrieval. *Pattern Recogn.* 42, 2 (Feb. 2009), 267–282.
- [55] Shan, C., Gong, S. and McOwan, P.W. 2009. Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study. *Image Vision Comput.* 27, 6 (May 2009), 803–816.
- [56] Shojaeilangari, S., Yau, W.-Y., Li, J. and Teoh, E.-K. 2014. Dynamic Facial Expression Analysis Based on Extended Spatio-temporal Histogram of Oriented Gradients. *Int. J. Biometrics.* 6, 1 (Mar. 2014), 33–52.
- [57] Thacker, N.A., Clark, A.F., Barron, J.L., Ross Beveridge, J., Courtney, P., Crum, W.R., Ramesh, V. and Clark, C. 2008. Performance Characterization in Computer Vision: A Guide to Best Practices. *Comput. Vis. Image Underst.* 109, 3 (Mar. 2008), 305–334.
- [58] Thome, N., Merad, D. and Miguët, S. 2008. Learning Articulated Appearance Models for Tracking Humans: A Spectral Graph Matching Approach. *Image Commun.* 23, 10 (Nov. 2008), 769–787.
- [59] Trémeau, A., Tominaga, S. and Plataniotis, K.N. 2008. Color in Image and Video

Processing: Most Recent Trends and Future Research Directions. *J. Image Video Process.* 2008, (Jan. 2008), 7:1–7:26.

[60] Turaga, P., Veeraraghavan, A. and Chellappa, R. 2009. Unsupervised View and Rate Invariant Clustering of Video Sequences. *Comput. Vis. Image Underst.* 113, 3 (Mar. 2009), 353–371.

[61] Vacca, J.R. and Vacca, J.R. 2013. *Computer and Information Security Handbook, Second Edition*. Morgan Kaufmann Publishers Inc.

[62] Vezzani, R., Baltieri, D. and Cucchiara, R. 2013. People Reidentification in Surveillance and Forensics: A Survey. *ACM Comput. Surv.* 46, 2 (Dec. 2013), 29:1–29:37.

[63] Vrigkas, M., Karavasili, V., Nikou, C. and Kakadiaris, I.A. 2014. Matching Mixtures of Curves for Human Action Recognition. *Comput. Vis. Image Underst.* 119, (Feb. 2014), 27–40.

[64] Walk, S., Schindler, K. and Schiele, B. 2010. Disparity Statistics for Pedestrian Detection: Combining Appearance, Motion and Stereo. *Proceedings of the 11th European Conference on Computer Vision: Part VI* (Berlin, Heidelberg, 2010), 182–195.

[65] Wang, C. and Li, Y. 2010. Combine Image Quality Fusion and Illumination Compensation for Video-based Face Recognition. *Neurocomput.* 73, 7-9 (Mar. 2010), 1478–1490.

[66] Wang, C., Wang, Y., Zhang, Z. and Wang, Y. 2014. Incremental Learning Patch-based Bag of Facial Words Representation for Face Recognition in Videos. *Multimedia Tools Appl.* 72, 3 (Oct. 2014), 2439–2467.

[67] Wang, X. 2013. Intelligent Multi-camera Video Surveillance: A Review. *Pattern Recogn. Lett.* 34, 1 (Jan. 2013), 3–19.

[68] Wolf, C., Lombardi, E., Mille, J., Celiktutan, O., Jiu, M., Dogan, E., Eren, G., Baccouche, M., Dellandréa, E., Bichot, C.-E., Garcia, C. and Sankur, B. 2014. Evaluation of Video Activity Localizations Integrating Quality and Quantity Measurements. *Comput. Vis. Image Underst.* 127, (Oct. 2014), 14–30.

[69] Wu, J., Hu, S. and Wang, Y. 2013. Adaptive Multifeature Visual Tracking in a Probability-hypothesis-density Filtering Framework. *Signal Process.* 93, 11 (Nov. 2013), 2915–2926.

- [70] Yilmaz, A., Javed, O. and Shah, M. 2006. Object Tracking: A Survey. *ACM Comput. Surv.* 38, 4 (Dec. 2006).
- [71] Zhang, Z., Huang, K., Wang, Y. and Li, M. 2013. View Independent Object Classification by Exploring Scene Consistency Information for Traffic Scene Surveillance. *Neurocomput.* 99, (Jan. 2013), 250–260.
- [72] Zhao, G., Huang, X., Taini, M., Li, S.Z. and Pietikäläinen, M. 2011. Facial Expression Recognition from Near-infrared Videos. *Image Vision Comput.* 29, 9 (Aug. 2011), 607–619.
- [73] Zhao, W., Chellappa, R., Phillips, P.J. and Rosenfeld, A. 2003. Face Recognition: A Literature Survey. *ACM Comput. Surv.* 35, 4 (Dec. 2003), 399–458.
- [74] Zheng, J., Ramírez, G.A. and Fuentes, O. 2010. Face Detection in Low-resolution Color Images. *Proceedings of the 7th International Conference on Image Analysis and Recognition - Volume Part I* (Berlin, Heidelberg, 2010), 454–463.
- [75] Zhou, H., Fei, M., Sadka, A., Zhang, Y. and Li, X. 2014. Adaptive Fusion of Particle Filtering and Spatio-temporal Motion Energy for Human Tracking. *Pattern Recogn.* 47, 11 (Nov. 2014), 3552–3567.
- [76] 2006. Author Index of Volume 27. *Pattern Recogn. Lett.* 27, 16 (Dec. 2006), 2013–2019.
- [77] 2006. Cumulative Contents of Volume 27. *Pattern Recogn. Lett.* 27, 16 (Dec. 2006), 2020–2027.

Crterios de incluso dos artigos:

- (a) Trabalhos que definirem tcnicas de reconhecimento facial.
- (b) Trabalhos que definem conceitos de processamento de imagens e de reconhecimento facial.
- (c) Trabalhos que se relacionem com o reconhecimento facial em ambientes reais, no controlados e com variao de iluminao.
- (d) Trabalhos que envolvem o reconhecimento facial utilizando cmeras de baixa resoluo.

Crterios de excluso dos artigos:

- (a) Trabalhos que no tenham relao com o reconhecimento facial.
- (b) Possveis trabalhos que utilizam tcnicas de reconhecimento facial, porm em ambientes controlados como, por exemplo, onde h boa iluminao e utilizam cmeras de alta resoluo.
- (c) Trabalhos que apresentem avaliao superficial, sem apresentar os mtodos e tcnicas utilizadas.

- (d) Trabalhos que realizam o reconhecimento facial utilizando modelos 3D.
 (e) Trabalhos que não tenham sido avaliados por seus pares.
 (f) Trabalhos que não envolvam o problema da variação de iluminação.

Métodos de avaliação:

Os artigos encontrados serão avaliados primeiramente através de seus títulos e resumos, e se necessário através da leitura de algumas passagens específicas do artigo como, por exemplo, os resultados e/ou a conclusão. Os artigos serão excluídos se atingirem pelo menos um dos critérios de exclusão preestabelecidos.

Lista de artigos com status de inclusão ou exclusão:

Artigo	Critérios de inclusão atendidos	Critérios de exclusão atendidos	Status
1.	(b)	(a)	Excluído
2.		(a)	Excluído
3.	(a)(c)		Incluído
4.	(a)(c)(d)		Incluído
5.	(a)(c)		Incluído
6.	(a)(c)(d)		Incluído
7.		(a)	Excluído
8.	(a)	(d)	Excluído
9.		(a)	Excluído
10.	(a)(c)		Incluído
11.	(a)		Incluído
12.	(b)		Incluído
13.		(a)	Excluído
14.		(a)	Excluído
15.		(a)	Excluído
16.		(a)	Excluído
17.		(a)	Excluído
18.		(a)	Excluído
19.		(a)	Excluído
20.	(b)(c)	(b)	Excluído
21.	(a)(b)	(f)	Excluído
22.		(a)	Excluído
23.		(a)	Excluído
24.		(a)	Excluído
25.		(a)	Excluído

26.		(a)	Excluído
27.		(a)	Excluído
28.		(a)	Excluído
29.		(a)	Excluído
30.	(a)(b)(c)	(f)	Excluído
31.	(a)(b)(c)	(f)	Excluído
32.		(a)	Excluído
33.		(a)	Excluído
34.		(a)	Excluído
35.		(a)	Excluído
36.	(c)	(c)(d)	Excluído
37.	(b)	(d)	Excluído
38.		(a)	Excluído
39.		(a)	Excluído
40.		(a)	Excluído
41.		(a)	Excluído
42.		(a)	Excluído
43.		(a)	Excluído
44.		(a)	Excluído
45.	(b)	(a)	Excluído
46.		(a)	Excluído
47.		(a)	Excluído
48.	(b)	(a)	Excluído
49.		(a)	Excluído
50.		(a)	Excluído
51.		(a)	Excluído
52.		(a)	Excluído
53.		(a)	Excluído
54.		(a)	Excluído
55.	(b)	(a)	Excluído
56.	(b)	(a)	Excluído
57.		(a)	Excluído
58.		(a)	Excluído
59.		(a)	Excluído
60.		(a)	Excluído
61.		(a)	Excluído

62.		(a)	Excluído
63.		(a)	Excluído
64.		(a)	Excluído
65.	(a)(b)(c)		Incluído
66.	(a)		Incluído
67.		(a)	Excluído
68.		(a)	Excluído
69.		(a)	Excluído
70.		(a)	Excluído
71.		(a)	Excluído
72.	(b)	(a)	Excluído
73.	(b)	Revisão da Literatura	Excluído
74.	(a)(b)(d)	(f)	Excluído
75.		(a)	Excluído
76.		(a)	Excluído
77.		(a)	Excluído

Apêndice C – Condução da Revisão na Base da IEEE

CONDUÇÃO DA REVISÃO SISTEMÁTICA

Fonte: IEEE Xplore (<http://ieeexplore.ieee.org/>)

Data de busca: 24/05/2015

Palavras-chave utilizadas: face recognition, face detection, low resolution, surveillance camera, illumination invariance, pose invariant.

String utilizada: (“face recognition” OR “face detection”) AND (“low resolution” OR “surveillance camera” OR “illumination invariance” OR “pose invariant”)

Total de artigos encontrados: 27

Total de artigos incluídos: 6

Organização dos artigos encontrados: Os artigos encontrados foram catalogados com o auxílio do software *Zotero*¹, e a utilização do padrão bibliográfico “*ACM SIGCHI Proceedings (2016)*”.

Lista de artigos encontrados:

[1] O. Arandjelović and R. Cipolla. 2008. Colour invariants for machine face recognition. *8th IEEE International Conference on Automatic Face Gesture Recognition, 2008. FG '08*, 1–8. <http://doi.org/10.1109/AFGR.2008.4813306>

[2] Hai-Wen Chen and M. McGurr. 2014. Improved color and intensity patch segmentation for human full-body and body-parts detection and tracking. *2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 361–368. <http://doi.org/10.1109/AVSS.2014.6918695>

[3] N.J. Chhasatia, C.U. Trivedi, and K.A. Shah. 2013. Performance evaluation of localized person based scene detection and retrieval in video. *2013 IEEE Second International Conference on Image Information Processing (ICIIP)*, 78–83. <http://doi.org/10.1109/ICIIP.2013.6707559>

[4] P.M. Corcoran, F. Nanu, S. Petrescu, and P. Bigioi. 2012. Real-time eye gaze tracking for gaming design and consumer electronics systems. *IEEE Transactions on Consumer Electronics* 58, 2, 347–355. <http://doi.org/10.1109/TCE.2012.6227433>

[5] P.R. Devarakota, M. Castillo-Franco, R. Ginhoux, B. Mirbach, S. Kater, and B. Ottersten. 2009. 3-D-Skeleton-Based Head Detection and Tracking Using Range Images. *IEEE Transactions on Vehicular Technology* 58, 8, 4064–4077.

¹ Website: <https://www.zotero.org/>

<http://doi.org/10.1109/TVT.2009.2020595>

[6] Chin-Shyurng Fahn and Chin-Sung Lo. 2010. A high-definition human face tracking system using the fusion of omni-directional and PTZ cameras mounted on a mobile robot. *2010 the 5th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 6–11. <http://doi.org/10.1109/ICIEA.2010.5514985>

[7] D. Gorodnichy and E. Granger. 2014. Target-based evaluation of face recognition technology for video surveillance applications. *2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, 110–117. <http://doi.org/10.1109/CIBIM.2014.7015451>

[8] M.A. Haque, K. Nasrollahi, and T.B. Moeslund. 2013. Real-time acquisition of high quality face sequences from an active pan-tilt-zoom camera. *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 443–448. <http://doi.org/10.1109/AVSS.2013.6636680>

[9] J. Harguess, Changbo Hu, and J.K. Aggarwal. 2009. Fusing face recognition from multiple cameras. *2009 Workshop on Applications of Computer Vision (WACV)*, 1–7. <http://doi.org/10.1109/WACV.2009.5403055>

[10] Jingu Heo, M. Savvides, and B. VijayaKumar. 2005. Performance Evaluation of Face Recognition using Visual and Thermal Imagery with Advanced Correlation Filters. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops*, 9–9. <http://doi.org/10.1109/CVPR.2005.508>

[11] J. Hirata, M. Morimoto, and K. Fujii. 2008. Estimating face direction from low resolution images. *Automation Congress, 2008. WAC 2008. World*, 1–6.

[12] Xiaolei Huang. 2008. Commentary Paper on “Video Surveillance for Biometrics: Long-Range Multi-biometric System.” *IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance, 2008. AVSS '08*, 183–183. <http://doi.org/10.1109/AVSS.2008.46>

[13] Y. Ishii, H. Hongo, K. Yamamoto, and Y. Niwa. 2004. Real-time face and head detection using four directional features. *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings*, 403–408. <http://doi.org/10.1109/AFGR.2004.1301566>

[14] Sung-Uk Jung and M.S. Nixon. 2011. Estimation of 3D head region using gait motion for

surveillance video. *4th International Conference on Imaging for Crime Detection and Prevention 2011 (ICDP 2011)*, 1–6. <http://doi.org/10.1049/ic.2011.0105>

[15] A. Leykin and R. Hammoud. 2008. Real-time estimation of human attention field in LWIR and color surveillance videos. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08*, 1–6. <http://doi.org/10.1109/CVPRW.2008.4563059>

[16] W. Louis and K.N. Plataniotis. 2010. Frontal face detection for surveillance purposes using dual Local Binary Patterns features. *2010 17th IEEE International Conference on Image Processing (ICIP)*, 3809–3812. <http://doi.org/10.1109/ICIP.2010.5653543>

[17] W. Louis, K.N. Plataniotis, and Yong Man Ro. 2010. Enhanced weakly trained frontal face detector for surveillance purposes. *2010 IEEE International Conference on Fuzzy Systems (FUZZ)*, 1–8. <http://doi.org/10.1109/FUZZY.2010.5584450>

[18] Yue Ming and Qiuqi Ruan. 2010. Face stereo matching and disparity calculation in binocular vision system. *2010 2nd International Conference on Industrial and Information Systems (IIS)*, 281–284. <http://doi.org/10.1109/INDUSIS.2010.5565855>

[19] Richa Mishra and Ravi Subban. 2014. Face detection for video summary using enhancement-based fusion strategy under varying illumination conditions. *2014 International Conference on Science Engineering and Management Research (ICSEMR)*, 1–8. <http://doi.org/10.1109/ICSEMR.2014.7043648>

[20] R. Mishra and R. Subban. 2014. Face detection for video summary using enhancement-based fusion strategy under varying illumination conditions. *2014 International Conference on Science Engineering and Management Research (ICSEMR)*, 1–8. <http://doi.org/10.1109/ICSEMR.2014.7043634>

[21] S.J.D. Prince, J.H. Elder, Y. Hou, and M. Sizinstev. 2005. Pre-Attentive Face Detection for Foveated Wide-Field Surveillance. *Seventh IEEE Workshops on Application of Computer Vision, 2005. WACV/MOTIONS '05 Volume 1*, 439–446. <http://doi.org/10.1109/ACVMOT.2005.95>

[22] R. Raghavendra, K.B. Raja, B. Yang, and C. Busch. 2013. A novel image fusion scheme for robust multiple face recognition with light-field camera. *2013 16th International Conference on Information Fusion (FUSION)*, 722–729.

[23] M. Rusinol, J. Lladós, and P. Dosch. 2007. Camera-Based Graphical Symbol Detection. *Ninth International Conference on Document Analysis and Recognition, 2007. ICDAR 2007*, 884–888. <http://doi.org/10.1109/ICDAR.2007.4377042>

[24] R. Thota, A. Kalyansundar, and Amit Kale. 2009. Modeling and tracking of faces in real-life illumination conditions. *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009*, 761–764. <http://doi.org/10.1109/ICASSP.2009.4959695>

[25] R. Valenti and T. Gevers. 2012. Accurate Eye Center Location through Invariant Isocentric Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 9, 1785–1798. <http://doi.org/10.1109/TPAMI.2011.251>

[26] T. Vijayan, N.A. Kumar, and K. Sivachandar. 2013. An improved low-resolution face tracker system using gradient logarithm field feature space and bi-cubic interpolation. *2013 International Conference on Human Computer Interactions (ICHCI)*, 1–7. <http://doi.org/10.1109/ICHCI-IEEE.2013.6887823>

[27] Zhifei Wang, Zhenjiang Miao, and Chao Zhang. 2008. Extraction of High-Resolution Face Image from Low-Resolution and Variant Illumination Video Sequences. *Congress on Image and Signal Processing, 2008. CISP '08*, 97–101. <http://doi.org/10.1109/CISP.2008.487>

Crterios de incluso dos artigos:

- (a) Trabalhos que definirem tcnicas de reconhecimento facial.
- (b) Trabalhos que definem conceitos de processamento de imagens e de reconhecimento facial.
- (c) Trabalhos que se relacionem com o reconhecimento facial em ambientes reais, no controlados e com variao de iluminao.
- (d) Trabalhos que envolvem o reconhecimento facial utilizando cmeras de baixa resoluo.

Crterios de excluso dos artigos:

- (a) Trabalhos que no tenham relao com o reconhecimento facial.
- (b) Possveis trabalhos que utilizam tcnicas de reconhecimento facial, porm em ambientes controlados como, por exemplo, onde h boa iluminao e utilizam cmeras de alta resoluo.
- (c) Trabalhos que apresentem avaliao superficial, sem apresentar os mtodos e tcnicas utilizadas.
- (d) Trabalhos que realizam o reconhecimento facial utilizando modelos 3D.
- (e) Trabalhos que no tenham sido avaliados por seus pares.
- (f) Trabalhos que no envolvam o problema da variao de iluminao.

Métodos de avaliação:

Os artigos encontrados serão avaliados primeiramente através de seus títulos e resumos, e se necessário através da leitura de algumas passagens específicas do artigo como, por exemplo, os resultados e/ou a conclusão. Os artigos serão excluídos se atingirem pelo menos um dos critérios de exclusão preestabelecidos.

Lista de artigos com status de inclusão ou exclusão:

Artigo	Critérios de inclusão atendidos	Critérios de exclusão atendidos	Status
1.	(a)(b)(c)		Incluído
2.	(b)	(a)	Excluído
3.	(b)(c)	(a)	Excluído
4.		(a)	Excluído
5.		(a)	Excluído
6.		(a)	Excluído
7.	(a)(b)(c)(d)	(f)	Excluído
8.	(a)(b)(c)(d)	(f)	Excluído
9.	(a)(b)(c)	(f)	Excluído
10.	(a)(b)		Incluído
11.		(a)	Excluído
12.		(a)	Excluído
13.		(a)	Excluído
14.		(a)	Excluído
15.		(a)	Excluído
16.	(a)(c)		Incluído
17.	(a)		Incluído
18.		(a)	Excluído
19.	(a)(b)(c)(d)		Incluído
20.	ARTIGO REPLICADO: IGUAL AO ARTIGO 19		
21.	(a)(d)	(f)	Excluído
22.	(a)(c)	(f)	Excluído
23.		(a)	Excluído
24.		(a)	Excluído
25.		(a)	Excluído
26.		(a)	Excluído
27.	(a)(b)(c)(d)		Incluído

Apêndice D – Termo de Consentimento Livre e Esclarecido

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Eu _____, RG _____, AUTORIZO, através do presente termo, os pesquisadores envolvidos no projeto de pesquisa intitulado "Detectando Presença Através de Identificação Facial sob Condições de Baixa Resolução e Luminosidade", o qual integra a linha de pesquisa de "Monitoramento de Presença" do projeto "Ensino e Monitoramento de Atividades Físicas via Técnicas de Inteligência Artificial" (decorrente de convênio firmado entre a Universidade de São Paulo, Faculdade Campo Limpo Paulista e Academia Central Kungfu-Wushu, conforme processo 2014.1.923.86.4, publicado no DOE 125(45), em 10/03/2015), bem como aos pesquisadores envolvidos em quaisquer sub-projetos a este ligados, a utilizar as imagens que se façam necessárias para fins científicos e didáticos (por exemplo, mas não limitados a livros, artigos científicos e material didático), sem quaisquer ônus a nenhuma das partes, obedecendo ao que está previsto nas Leis que resguardam os direitos das crianças e adolescentes (Estatuto da Criança e do Adolescente – ECA, Lei N.º 8.069/1990), dos idosos (Estatuto do Idoso, Lei N.º 10.741/2003) e das pessoas com deficiência (Decreto No 3.298/1999, alterado pelo Decreto No 5.296/2004).

O(a) participante da pesquisa contribuirá para acrescentar à literatura dados referentes ao tema processamento de imagens, mais especificamente a área de reconhecimento facial. Ainda, estará contribuindo para o desenvolvimento de novas tecnologias, que por sua vez, podem se reverter para o seu bem em outros cenários.

O risco psicológico é mínimo pois serão utilizadas imagens já existentes, obtidas das câmeras de segurança já instaladas na academia.

Faço isso sabendo que, em nenhum momento, serão divulgadas informações, para além da imagem em si, que permitam a identificação do participante da pesquisa.

Caso o presente participante seja menor de idade o Termo de Assentimento deverá ser assinado pelo seu responsável.

O presente termo será assinado em duas vias, sendo uma para o participante e outra para o pesquisador responsável.

_____, _____ de _____ de 20

Nome do participante da pesquisa:

RG: CPF:

Assinatura:

Nome do membro da equipe da pesquisa:

RG: CPF:

Assinatura:

Nome do representante legal:

RG: CPF:

Parentesco:

Assinatura:

Pesquisador: Kelvin Salton do Prado
Telefone: (11) 2305-5623
Email: kelvinpfw@hotmail.com

Comitê de Ética em Pesquisa Envolvendo Seres Humanos - EACH
Telefone: (11) 3091-1046
Email: cep-each@usp.br

Apêndice E – Termo de Assentimento

TERMO DE ASSENTIMENTO

O assentimento informado para a criança/adolescente não substitui a necessidade de consentimento informado dos pais ou responsáveis. O assentimento assinado pela criança/adolescente demonstra a sua cooperação com a pesquisa.

Projeto de Pesquisa: Detectando Presença Através de Identificação Facial sob Condições de Baixa Resolução e Luminosidade.

Local da Pesquisa: Academia Central Kungfu-Wushu.

Informações:

Você está sendo convidado(a) a participar de uma pesquisa, com o objetivo de identificar a melhor técnica de reconhecimento facial que se adapte ao ambiente real e não controlado da academia. A pesquisa visa a apresentar as técnicas que melhor se adaptam ao ambiente proposto, e se possível propor novas abordagens que possam de certa forma melhorar as técnicas de reconhecimento facial já existentes. Para isto serão coletadas imagens e vídeos da academia alvo do estudo para a realização dos testes e aplicação dos algoritmos e técnicas de reconhecimento facial. O participante do projeto estará contribuindo para o desenvolvimento da área de pesquisa de processamento de imagens, mais precisamente para a área de reconhecimento facial, ainda, estará contribuindo para o desenvolvimento de novas tecnologias, que por sua vez, podem se reverter para o seu bem em outros cenários. Como consta no termo de consentimento livre e esclarecido, as imagens cedidas serão utilizadas nas publicações referentes ao presente projeto e demais projetos ligados a este. Com isto, acredita-se que o risco psicológico é mínimo pois serão utilizadas imagens já existentes, obtidas das câmeras de segurança da academia.

Concordando com este termo o participante autoriza a utilização das imagens e vídeos, cedidos pela academia Kungfu-Wushu, os quais serão utilizados no desenvolvimento do projeto de pesquisa intitulado “Detectando Presença Através de Identificação Facial sob Condições de Baixa Resolução e Luminosidade”.

Declaração de Assentimento

Li e entendi os objetivos da pesquisa descritos neste documento. Entendo que sou livre para aceitar ou recusar, e que posso interromper minha participação a qualquer momento. Eu concordo que os dados para o estudo sejam coletados para o propósito descrito neste documento.

Eu _____, inscrito no RG _____,
 AUTORIZO o aluno _____, inscrito no RG _____,
 a participar do projeto de pesquisa intitulado “Detectando Presença Através de Identificação Facial sob Condições de Baixa Resolução e Luminosidade”, o qual integra a linha de pesquisa de “Monitoramento de Presença” do projeto “Ensino e Monitoramento de Atividades Físicas via Técnicas de Inteligência Artificial” (decorrente de convênio firmado entre a Universidade de São Paulo, Faculdade Campo Limpo Paulista e Academia Central Kungfu-Wushu, conforme processo 2014.1.923.86.4, publicado no DOE 125(45), em 10/03/2015), bem como em quaisquer sub-projetos a este ligados, sem quaisquer ônus a nenhuma das partes, tendo lido e entendido os benefícios e riscos do projeto.

O presente termo será assinado em duas vias, sendo uma para o participante e outra para o pesquisador responsável.

_____, ____ de _____ de 20

Nome do participante da pesquisa:

RG: CPF:

Assinatura:

Nome do representante legal:

RG: CPF:

Parentesco:

Assinatura:

Nome do membro da equipe da pesquisa:

RG: CPF:

Assinatura:

Pesquisador: Kelvin Salton do Prado
 Telefone: (11) 2305-5623
 Email: kelvinpww@hotmail.com

Comitê de Ética em Pesquisa Envolvendo Seres Humanos - EACH
 Telefone: (11) 3091-1046
 Email: cep-each@usp.br

Apêndice F – Resultados dos Testes com os Diversos Métodos de Interpolação

Resultados testes utilizando diferentes métodos de redimensionamento de imagens								
Base de Treinamento	Base de Teste	Algoritmo	Parâmetros	Resolução	Parâmetro Interpolação	Faces Reconhecidas	Faces Não Reconhecidas	Não são Faces
Base 1	Vídeo 2	Eigenface	Padrão	100x100	INTER_CUBIC	6	571	57
Base 1	Vídeo 2	Eigenface	Padrão	100x100	INTER_NEAREST	6	571	57
Base 1	Vídeo 2	Eigenface	Padrão	100x100	INTER_LINEAR	7	570	57
Base 1	Vídeo 2	Eigenface	Padrão	100x100	INTER_AREA	7	570	57
Base 1	Vídeo 2	Eigenface	Padrão	100x100	INTER_LANCZOS4	7	570	57
Base 1	Vídeo 2	Fisherface	Padrão	100x100	INTER_CUBIC	6	571	57
Base 1	Vídeo 2	Fisherface	Padrão	100x100	INTER_NEAREST	5	572	57
Base 1	Vídeo 2	Fisherface	Padrão	100x100	INTER_LINEAR	7	570	57
Base 1	Vídeo 2	Fisherface	Padrão	100x100	INTER_AREA	7	570	57
Base 1	Vídeo 2	Fisherface	Padrão	100x100	INTER_LANCZOS4	7	570	57
Base 1	Vídeo 2	LBPH	Padrão	100x100	INTER_CUBIC	10	567	57
Base 1	Vídeo 2	LBPH	Padrão	100x100	INTER_NEAREST	1	576	57
Base 1	Vídeo 2	LBPH	Padrão	100x100	INTER_LINEAR	14	563	57
Base 1	Vídeo 2	LBPH	Padrão	100x100	INTER_AREA	0	577	57
Base 1	Vídeo 2	LBPH	Padrão	100x100	INTER_LANCZOS4	14	563	57
Base 1	Vídeo 2	SIFT	Padrão	100x100	INTER_CUBIC	11	566	57
Base 1	Vídeo 2	SIFT	Padrão	100x100	INTER_NEAREST	7	570	57
Base 1	Vídeo 2	SIFT	Padrão	100x100	INTER_LINEAR	12	565	57
Base 1	Vídeo 2	SIFT	Padrão	100x100	INTER_AREA	13	564	57
Base 1	Vídeo 2	SIFT	Padrão	100x100	INTER_LANCZOS4	11	566	57
Base 1	Vídeo 2	SURF	Padrão	100x100	INTER_CUBIC	22	555	57
Base 1	Vídeo 2	SURF	Padrão	100x100	INTER_NEAREST	16	561	57
Base 1	Vídeo 2	SURF	Padrão	100x100	INTER_LINEAR	10	567	57
Base 1	Vídeo 2	SURF	Padrão	100x100	INTER_AREA	13	564	57
Base 1	Vídeo 2	SURF	Padrão	100x100	INTER_LANCZOS4	23	554	57
Base 1	Vídeo 2	Eigenface	Padrão	200x200	INTER_CUBIC	7	570	57
Base 1	Vídeo 2	Eigenface	Padrão	200x200	INTER_NEAREST	6	571	57
Base 1	Vídeo 2	Eigenface	Padrão	200x200	INTER_LINEAR	7	570	57
Base 1	Vídeo 2	Eigenface	Padrão	200x200	INTER_AREA	7	570	57
Base 1	Vídeo 2	Eigenface	Padrão	200x200	INTER_LANCZOS4	6	571	57
Base 1	Vídeo 2	Fisherface	Padrão	200x200	INTER_CUBIC	6	571	57
Base 1	Vídeo 2	Fisherface	Padrão	200x200	INTER_NEAREST	6	571	57
Base 1	Vídeo 2	Fisherface	Padrão	200x200	INTER_LINEAR	6	571	57
Base 1	Vídeo 2	Fisherface	Padrão	200x200	INTER_AREA	7	570	57
Base 1	Vídeo 2	Fisherface	Padrão	200x200	INTER_LANCZOS4	6	571	57
Base 1	Vídeo 2	LBPH	Padrão	200x200	INTER_CUBIC	1	576	57
Base 1	Vídeo 2	LBPH	Padrão	200x200	INTER_NEAREST	0	577	57
Base 1	Vídeo 2	LBPH	Padrão	200x200	INTER_LINEAR	1	576	57
Base 1	Vídeo 2	LBPH	Padrão	200x200	INTER_AREA	0	577	57
Base 1	Vídeo 2	LBPH	Padrão	200x200	INTER_LANCZOS4	0	577	57
Base 1	Vídeo 2	SIFT	Padrão	200x200	INTER_CUBIC	4	573	57
Base 1	Vídeo 2	SIFT	Padrão	200x200	INTER_NEAREST	18	559	57
Base 1	Vídeo 2	SIFT	Padrão	200x200	INTER_LINEAR	11	566	57
Base 1	Vídeo 2	SIFT	Padrão	200x200	INTER_AREA	17	560	57
Base 1	Vídeo 2	SIFT	Padrão	200x200	INTER_LANCZOS4	10	567	57
Base 1	Vídeo 2	SURF	Padrão	200x200	INTER_CUBIC	19	558	57
Base 1	Vídeo 2	SURF	Padrão	200x200	INTER_NEAREST	19	558	57
Base 1	Vídeo 2	SURF	Padrão	200x200	INTER_LINEAR	28	549	57
Base 1	Vídeo 2	SURF	Padrão	200x200	INTER_AREA	23	554	57
Base 1	Vídeo 2	SURF	Padrão	200x200	INTER_LANCZOS4	15	562	57
Base 6	Vídeo 2	Eigenface	Padrão	100x100	INTER_CUBIC	11	566	57
Base 6	Vídeo 2	Eigenface	Padrão	100x100	INTER_NEAREST	12	565	57
Base 6	Vídeo 2	Eigenface	Padrão	100x100	INTER_LINEAR	13	564	57
Base 6	Vídeo 2	Eigenface	Padrão	100x100	INTER_AREA	12	565	57
Base 6	Vídeo 2	Eigenface	Padrão	100x100	INTER_LANCZOS4	11	566	57
Base 6	Vídeo 2	Fisherface	Padrão	100x100	INTER_CUBIC	93	484	57
Base 6	Vídeo 2	Fisherface	Padrão	100x100	INTER_NEAREST	51	526	57
Base 6	Vídeo 2	Fisherface	Padrão	100x100	INTER_LINEAR	84	493	57
Base 6	Vídeo 2	Fisherface	Padrão	100x100	INTER_AREA	106	471	57
Base 6	Vídeo 2	Fisherface	Padrão	100x100	INTER_LANCZOS4	91	486	57
Base 6	Vídeo 2	LBPH	Padrão	100x100	INTER_CUBIC	9	568	57
Base 6	Vídeo 2	LBPH	Padrão	100x100	INTER_NEAREST	4	573	57
Base 6	Vídeo 2	LBPH	Padrão	100x100	INTER_LINEAR	16	561	57
Base 6	Vídeo 2	LBPH	Padrão	100x100	INTER_AREA	15	562	57
Base 6	Vídeo 2	LBPH	Padrão	100x100	INTER_LANCZOS4	14	563	57
Base 6	Vídeo 2	SIFT	Padrão	100x100	INTER_CUBIC	26	551	57
Base 6	Vídeo 2	SIFT	Padrão	100x100	INTER_NEAREST	35	542	57
Base 6	Vídeo 2	SIFT	Padrão	100x100	INTER_LINEAR	24	553	57
Base 6	Vídeo 2	SIFT	Padrão	100x100	INTER_AREA	15	562	57
Base 6	Vídeo 2	SIFT	Padrão	100x100	INTER_LANCZOS4	32	545	57

Base 6	Video 2	SURF	Padrão	100x100	INTER_CUBIC	21	556	57
Base 6	Video 2	SURF	Padrão	100x100	INTER_NEAREST	12	565	57
Base 6	Video 2	SURF	Padrão	100x100	INTER_LINEAR	15	562	57
Base 6	Video 2	SURF	Padrão	100x100	INTER_AREA	21	556	57
Base 6	Video 2	SURF	Padrão	100x100	INTER_LANCZOS4	22	555	57
Base 6	Video 2	Eigenface	Padrão	200x200	INTER_CUBIC	12	565	57
Base 6	Video 2	Eigenface	Padrão	200x200	INTER_NEAREST	12	565	57
Base 6	Video 2	Eigenface	Padrão	200x200	INTER_LINEAR	12	565	57
Base 6	Video 2	Eigenface	Padrão	200x200	INTER_AREA	11	566	57
Base 6	Video 2	Eigenface	Padrão	200x200	INTER_LANCZOS4	11	566	57
Base 6	Video 2	Fisherface	Padrão	200x200	INTER_CUBIC	111	466	57
Base 6	Video 2	Fisherface	Padrão	200x200	INTER_NEAREST	56	521	57
Base 6	Video 2	Fisherface	Padrão	200x200	INTER_LINEAR	87	490	57
Base 6	Video 2	Fisherface	Padrão	200x200	INTER_AREA	64	513	57
Base 6	Video 2	Fisherface	Padrão	200x200	INTER_LANCZOS4	79	498	57
Base 6	Video 2	LBPH	Padrão	200x200	INTER_CUBIC	0	577	57
Base 6	Video 2	LBPH	Padrão	200x200	INTER_NEAREST	0	577	57
Base 6	Video 2	LBPH	Padrão	200x200	INTER_LINEAR	3	574	57
Base 6	Video 2	LBPH	Padrão	200x200	INTER_AREA	0	577	57
Base 6	Video 2	LBPH	Padrão	200x200	INTER_LANCZOS4	0	577	57
Base 6	Video 2	SIFT	Padrão	200x200	INTER_CUBIC	29	548	57
Base 6	Video 2	SIFT	Padrão	200x200	INTER_NEAREST	42	535	57
Base 6	Video 2	SIFT	Padrão	200x200	INTER_LINEAR	25	552	57
Base 6	Video 2	SIFT	Padrão	200x200	INTER_AREA	37	540	57
Base 6	Video 2	SIFT	Padrão	200x200	INTER_LANCZOS4	51	526	57
Base 6	Video 2	SURF	Padrão	200x200	INTER_CUBIC	27	550	57
Base 6	Video 2	SURF	Padrão	200x200	INTER_NEAREST	23	554	57
Base 6	Video 2	SURF	Padrão	200x200	INTER_LINEAR	33	544	57
Base 6	Video 2	SURF	Padrão	200x200	INTER_AREA	33	544	57
Base 6	Video 2	SURF	Padrão	200x200	INTER_LANCZOS4	36	541	57
Base 4	Video 2	Eigenface	Padrão	100x100	INTER_CUBIC	7	570	57
Base 4	Video 2	Eigenface	Padrão	100x100	INTER_NEAREST	6	571	57
Base 4	Video 2	Eigenface	Padrão	100x100	INTER_LINEAR	7	570	57
Base 4	Video 2	Eigenface	Padrão	100x100	INTER_AREA	7	570	57
Base 4	Video 2	Eigenface	Padrão	100x100	INTER_LANCZOS4	7	570	57
Base 4	Video 2	Fisherface	Padrão	100x100	INTER_CUBIC	11	566	57
Base 4	Video 2	Fisherface	Padrão	100x100	INTER_NEAREST	8	569	57
Base 4	Video 2	Fisherface	Padrão	100x100	INTER_LINEAR	13	564	57
Base 4	Video 2	Fisherface	Padrão	100x100	INTER_AREA	13	564	57
Base 4	Video 2	Fisherface	Padrão	100x100	INTER_LANCZOS4	11	566	57
Base 4	Video 2	LBPH	Padrão	100x100	INTER_CUBIC	11	566	57
Base 4	Video 2	LBPH	Padrão	100x100	INTER_NEAREST	0	577	57
Base 4	Video 2	LBPH	Padrão	100x100	INTER_LINEAR	14	563	57
Base 4	Video 2	LBPH	Padrão	100x100	INTER_AREA	0	577	57
Base 4	Video 2	LBPH	Padrão	100x100	INTER_LANCZOS4	15	562	57
Base 4	Video 2	SIFT	Padrão	100x100	INTER_CUBIC	12	565	57
Base 4	Video 2	SIFT	Padrão	100x100	INTER_NEAREST	16	561	57
Base 4	Video 2	SIFT	Padrão	100x100	INTER_LINEAR	21	556	57
Base 4	Video 2	SIFT	Padrão	100x100	INTER_AREA	11	566	57
Base 4	Video 2	SIFT	Padrão	100x100	INTER_LANCZOS4	19	558	57
Base 4	Video 2	SURF	Padrão	100x100	INTER_CUBIC	17	560	57
Base 4	Video 2	SURF	Padrão	100x100	INTER_NEAREST	17	560	57
Base 4	Video 2	SURF	Padrão	100x100	INTER_LINEAR	17	560	57
Base 4	Video 2	SURF	Padrão	100x100	INTER_AREA	12	565	57
Base 4	Video 2	SURF	Padrão	100x100	INTER_LANCZOS4	23	554	57
Base 4	Video 2	Eigenface	Padrão	200x200	INTER_CUBIC	7	570	57
Base 4	Video 2	Eigenface	Padrão	200x200	INTER_NEAREST	7	570	57
Base 4	Video 2	Eigenface	Padrão	200x200	INTER_LINEAR	7	570	57
Base 4	Video 2	Eigenface	Padrão	200x200	INTER_AREA	7	570	57
Base 4	Video 2	Eigenface	Padrão	200x200	INTER_LANCZOS4	7	570	57
Base 4	Video 2	Fisherface	Padrão	200x200	INTER_CUBIC	11	566	57
Base 4	Video 2	Fisherface	Padrão	200x200	INTER_NEAREST	14	563	57
Base 4	Video 2	Fisherface	Padrão	200x200	INTER_LINEAR	14	563	57
Base 4	Video 2	Fisherface	Padrão	200x200	INTER_AREA	11	566	57
Base 4	Video 2	Fisherface	Padrão	200x200	INTER_LANCZOS4	11	566	57
Base 4	Video 2	LBPH	Padrão	200x200	INTER_CUBIC	3	574	57
Base 4	Video 2	LBPH	Padrão	200x200	INTER_NEAREST	1	576	57
Base 4	Video 2	LBPH	Padrão	200x200	INTER_LINEAR	3	574	57
Base 4	Video 2	LBPH	Padrão	200x200	INTER_AREA	0	577	57
Base 4	Video 2	LBPH	Padrão	200x200	INTER_LANCZOS4	1	576	57
Base 4	Video 2	SIFT	Padrão	200x200	INTER_CUBIC	12	565	57
Base 4	Video 2	SIFT	Padrão	200x200	INTER_NEAREST	31	546	57
Base 4	Video 2	SIFT	Padrão	200x200	INTER_LINEAR	9	568	57

Base 4	Video 2	SIFT	Padrão	200x200	INTER_AREA	11	566	57
Base 4	Video 2	SIFT	Padrão	200x200	INTER_LANCZOS4	13	564	57
Base 4	Video 2	SURF	Padrão	200x200	INTER_CUBIC	21	556	57
Base 4	Video 2	SURF	Padrão	200x200	INTER_NEAREST	19	558	57
Base 4	Video 2	SURF	Padrão	200x200	INTER_LINEAR	20	557	57
Base 4	Video 2	SURF	Padrão	200x200	INTER_AREA	41	536	57
Base 4	Video 2	SURF	Padrão	200x200	INTER_LANCZOS4	33	544	57