

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E  
INFORMÁTICA INDUSTRIAL**

**FLÁVIO DAS CHAGAS PRODOSSIMO**

**METODOLOGIA PARA A CAPTURA, DETECÇÃO E  
NORMALIZAÇÃO DE IMAGENS FACIAIS**

**DISSERTAÇÃO**

**CURITIBA  
2013**

**FLÁVIO DAS CHAGAS PRODOSSIMO**

**METODOLOGIA PARA A CAPTURA, DETECÇÃO E  
NORMALIZAÇÃO DE IMAGENS FACIAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial da Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção do grau de “Mestre em Ciências” – Área de Concentração: Engenharia da Computação.

Orientador: Prof. Dr. Heitor Silvério Lopes

**CURITIBA  
2013**

---

Dados Internacionais de Catalogação na Publicação

---

- P964 Prodossimo, Flávio das Chagas  
Metodologia para a captura, detecção e normalização de imagens faciais / Flávio das Chagas Prodossimo. – 2013.  
107 f. : il. ; 30 cm
- Orientador: Heitor Silvério Lopes.  
Dissertação (Mestrado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial. Curitiba, 2013.  
Bibliografia: f. 82-85
1. Reconhecimento de rosto humano (Computação). 2. Percepção facial. 3. Sistemas de reconhecimento de padrões. 4. Processamento de imagens – Matemática. 5. Normalização. 6. Visão por computador. 7. Engenharia elétrica – Dissertações. I. Lopes, Heitor Silvério, orient. II. Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial. III. Título.

CDD (22. ed.) 621.3

**Título da Dissertação Nº. 634**

## **“Metodologia para a Captura, Detecção e Normalização de Imagens Faciais”**

por

**Flávio das Chagas Prodossimo**

Esta dissertação foi apresentada como requisito parcial à obtenção do grau de MESTRE EM CIÊNCIAS – Área de Concentração: Engenharia de Computação, pelo Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial – CPGEI – da Universidade Tecnológica Federal do Paraná – UTFPR, às 8h do dia 29 de maio de 2013. O trabalho foi aprovado pela Banca Examinadora, composta pelos doutores:

---

Prof. Heitor Silvério Lopes, Dr.  
(Presidente – UTFPR)

---

Prof. Luiz Antônio Pereira Neves, Dr.  
(UFPR)

---

Prof<sup>a</sup> Leyza Elmeri Baldo Dorini, Dr<sup>a</sup>.  
(UTFPR)

Visto da coordenação:

---

Prof. Ricardo Lüders, Dr.  
(Coordenador do CPGEI)

Dedico esta dissertação à Deus e aos meus pais, por sempre estarem presentes na minha vida.

## **AGRADECIMENTOS**

Primeiramente quero agradecer aos meus pais, Raquel e Valdomiro, pelo apoio em todos os momentos da minha vida. Aguentando meus momentos de mau humor durante os estudos e trabalhos.

Ao meu orientador, Prof. Heitor, pela orientação e pelas diversas discussões durante a realização deste trabalho.

Aos colegas do laboratório de Bioinformática, principalmente ao Chidam e ao Marlon, pelas várias discussões sobre o trabalho. Novamente ao Chidam pelas lições de vida passadas durante horas de conversas no laboratório.

Ao meu primo Lucas, minha irmã Claudia e meu cunhado Aurélio, por me aguentarem em todo o tempo de Mestrado e também pelos momentos de descontração. Também a Claudia pelas correções e dicas para esta dissertação.

Agradeço, também, aos meus amigos de Guarapuava que me apoiaram desde antes de iniciar o Mestrado, em especial aos Professores Richard e Carol.

A minha amiga Chris e minha prima Dé por 'ouvirem' e aguentarem meus desabafos de stress durante esta jornada.

Agradeço a todos os participantes da captura de imagens, que sem eles este trabalho não seria possível. A CAPES pela concessão da bolsa durante grande parte do trabalho. E aos colegas da CELEPAR que, mesmo nos momentos de conclusão do trabalho, me deram apoio.

Por fim, a todos que estiveram presentes na minha vida durante todo o mestrado.

”Aprender uma coisa significa entrar em contato com um mundo do qual não se tem a menor ideia. É preciso ser humilde para aprender.”  
(Paulo Coelho)

E você sempre tem algo novo para aprender...

## RESUMO

PRODOSSIMO, Flávio das Chagas. METODOLOGIA PARA A CAPTURA, DETECÇÃO E NORMALIZAÇÃO DE IMAGENS FACIAIS. 107 f. Dissertação – Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2013.

O reconhecimento facial está se tornando uma tarefa comum com a evolução da tecnologia da informação. Este artefato pode ser utilizado na área de segurança, controlando acesso a lugares restritos, identificando pessoas que tenham cometido atos ilícitos, entre outros. Executar o reconhecimento facial é uma tarefa complexa e, para completar este processo, são implementadas etapas que compreendem: a captura de imagens faciais, a detecção de regiões de interesse, a normalização facial, a extração de características e o reconhecimento em si. Dentre estas, as três primeiras são tratadas neste trabalho, que tem como objetivo principal a normalização automática de faces. Tanto para a captura de imagens quanto para a normalização frontal existem normas internacionais que padronizam o procedimento de execução destas tarefas e que foram utilizadas neste trabalho. Além disto, algumas normas foram adaptadas para a construção de uma base de imagens faciais com o objetivo de auxiliar o processo de reconhecimento facial. Também foi criada uma nova metodologia para normalização de imagens faciais laterais, baseando-se nas normas da normalização frontal. Foram implementadas normalização semiautomática frontal, semiautomática lateral e automática lateral. Para a execução da normalização facial automática são necessários dois pontos de controle, os dois olhos, o que torna indispensável a execução da etapa de detecção de regiões de interesse. Neste trabalho, foram comparadas duas metodologias semelhantes para detecção. Primeiramente foi detectada uma região contendo ambos os olhos e, em seguida, dentro desta região, foram detectados cada um dos olhos de forma mais precisa. Para as duas metodologias foram utilizadas técnicas de processamento de imagens e reconhecimento de padrões. A primeira metodologia utiliza como filtro o *Haar-Like Features* em conjunto com a técnica de reconhecimento de padrões *Adaptive Boosting*. Sendo que as técnicas equivalentes no segundo algoritmo foram o *Local Binary Pattern* e o *Support Vector Machines*, respectivamente. Na segunda metodologia também foi utilizado um algoritmo de otimização de busca baseado em vizinhança, o *Variable Neighborhood Search*. Os estudos resultaram em uma base com 3726 imagens, mais uma base normalizada frontal com 966 imagens e uma normalizada lateral com 276 imagens. A detecção de olhos resultou, nos melhores testes, em aproximadamente 99% de precisão para a primeira metodologia e 95% para a segunda, sendo que em todos os testes a primeira foi a mais rápida. Com o desenvolvimento de trabalhos futuros pretende-se: tornar públicas as bases de imagens, melhorar a porcentagem de acerto e velocidade de processamento para todos os testes e melhorar a normalização, implementando a normalização de plano de fundo e também de iluminação.

**Palavras-chave:** Base de Imagens Faciais, Normalização Facial, Detecção de ROIs, Reconhecimento Facial



## ABSTRACT

PRODOSSIMO, Flávio das Chagas. METHODOLOGY FOR CAPTURE, DETECTION AND NORMALIZATION OF FACIAL IMAGES. 107 f. Dissertação – Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2013.

With the evolution of information technology Facial recognition is becoming a common task. This artifact can be used in security, controlling access to restricted places and identifying persons, for example. Facial recognition is a complex task, and it's divided into some process, comprising: facial images capture, detection of regions of interest, facial normalization, feature extraction and recognition itself. Among these, the first three are treated in this work, which has as its main objective the automatic normalization of faces. For the capture of images and for the image normalization there are international standards that standardize the procedure for implementing these tasks and which were used in this work. In addition to following these rules, other standardizations have been developed to build a database of facial images in order to assist the process of face recognition. A new methodology for normalization of profile faces, based on the rules of frontal normalization. Some ways of normalization were implemented: frontal semiautomatic, lateral semiautomatic and automatic frontal. For the execution of frontal automatic normalization we need two points of interest, the two eyes, which makes it a necessary step to execute the detection regions of interest. In this work, we compared two similar methods for detecting. Where was first detected a region containing both eyes and then, within this region were detected each eye more accurately. For the two methodologies were used techniques of image processing and pattern recognition. The first method based on the Viola and Jones algorithm, the filter uses as Haar-like Features with the technique of pattern recognition Adaptive Boosting. Where the second algorithm equivalent techniques were Local Binary Pattern and Support Vector Machines, respectively. In the second algorithm was also used an optimization algorithm based on neighborhood search, the Variable Neighborhood Search. This studies resulted in a database with 3726 images, a frontal normalized database with 966 images and a database with face's profile normalized with 276 images. The eye detection resulted in better tests, about 99 % accuracy for the first method and 95 % for the second, and in all tests the first algorithm was the fastest. With the development of future work we have: make public the images database, improve the percentage of accuracy and processing speed for all tests and improve the normalization by implementing the normalization of the background and also lighting.

**Keywords:** Face Images Database, Face Normalization, ROIs Detection, Facial Recognition

## LISTA DE FIGURAS

FIGURA 1	– Fluxograma das Etapas de um Sistema de Visão Computacional	19
FIGURA 2	– Procedimento do LBP	20
FIGURA 3	– Modelos de <i>Haar-like Features</i>	22
FIGURA 4	– Processo de extração de características do <i>Haar-like</i>	23
FIGURA 5	– Exemplos de Elementos Estruturantes	24
FIGURA 6	– Exemplo do Processo de Erosão	25
FIGURA 7	– Exemplo do Processo de Dilatação	26
FIGURA 8	– Exemplo do Processo de Abertura	27
FIGURA 9	– Exemplo do Processo de Fechamento	27
FIGURA 10	– Exemplo de Segmentação	28
FIGURA 11	– Etapas de um algoritmo de Reconhecimento de Padrões	29
FIGURA 12	– Exemplo de classes separadas por um hiperplano	30
FIGURA 13	– Pseudo-código do <i>Adaptive Boosting</i> .	31
FIGURA 14	– Pseudo-código do <i>Variable Neighborhood Search</i> .	32
FIGURA 15	– Imagem Frontal Centralizada	35
FIGURA 16	– Exemplo de Imagens da Base FERET	38
FIGURA 17	– Região dos Olhos	39
FIGURA 18	– Imagem Frontal Normalizada Canônica	42
FIGURA 19	– Processo de Normalização de uma Imagem Facial	42
FIGURA 20	– Distâncias e Posições do Equipamento para Captura de Imagens	48
FIGURA 21	– Efeitos das diferentes formas de Iluminação	49
FIGURA 22	– Eixos de rotação e translação da face	51
FIGURA 23	– Etapas para detecção dos olhos	54
FIGURA 24	– Exemplos de regiões e não-regiões de olhos	55
FIGURA 25	– Exemplos de olhos e não-olhos	55
FIGURA 26	– Medidas da Imagem Normalizada	57
FIGURA 27	– Processo de Rotação dos Olhos	58
FIGURA 28	– Processo de Redimensionamento da Face	59
FIGURA 29	– Medidas da Face Lateral Normalizada	61
FIGURA 30	– Trago - Região da Orelha	61
FIGURA 31	– Seleção dos pontos de referência para normalização frontal	63
FIGURA 32	– Seleção dos pontos de referência para normalização lateral	63
FIGURA 33	– Processo automático de normalização	64
FIGURA 34	– Exemplos de seqüências de imagens dos testes preliminares	66
FIGURA 35	– Ambiente da captura de imagens	67
FIGURA 36	– Exemplos de imagens excluídas da base final	68
FIGURA 37	– Exemplos de imagens da base facial construída	69
FIGURA 38	– Exemplos das imagens utilizadas nos experimentos de detecção de olhos	71
FIGURA 39	– Gráfico de comparação entre os acertos dos algoritmos	72
FIGURA 40	– Regiões detectadas erroneamente	73
FIGURA 41	– Etapas da normalização frontal, com ampliação da imagem	74
FIGURA 42	– Etapas da normalização frontal, com redução da imagem	74

FIGURA 43	– Exemplos de imagens frontais normalizadas .....	75
FIGURA 44	– Etapas da Normalização de Plano de Fundo .....	76
FIGURA 45	– Exemplos da Normalização de Plano de Fundo .....	77
FIGURA 46	– Etapas da normalização lateral, com rotação para cima .....	77
FIGURA 47	– Etapas da normalização lateral, com rotação para baixo .....	77
FIGURA 48	– Exemplos de imagens laterais normalizadas .....	78
FIGURA 49	– Termo de Autorização .....	86

## LISTA DE TABELAS

TABELA 1	– Configurações da câmera e da imagem aconselhadas pelas normas internacionais. ....	37
TABELA 2	– Comparativo de Bases Faciais .....	37
TABELA 3	– Configurações da câmera e da imagem utilizadas. ....	47
TABELA 4	– Sequência de Captura das Imagens para cada cenário .....	52
TABELA 5	– Codificação dos Arquivos de Imagem .....	53
TABELA 6	– Codificação da Faixa Etária .....	53
TABELA 7	– Exemplo Codificação dos Arquivos de Imagem .....	69
TABELA 8	– Resumo das características da Base de Imagens Faciais Construída ....	70
TABELA 9	– Resultados da Detecção de Olhos com o algoritmo HA .....	71
TABELA 10	– Resultados da Detecção de Olhos com o algoritmo LS .....	71

## LISTA DE SIGLAS

RF	Reconhecimento Facial
ROI	<i>Region of Interest</i>
LBP	<i>Local Binary Pattern</i>
SVM	<i>Support Vector Machines</i>
Adaboost	<i>Adaptative Boosting</i>
VNS	<i>Variable Neighborhood Search</i>
HSI	Hue-Saturation-Intensity
NIST	<i>National Institute of Standards and Technology</i>
RGB	<i>Red Green Blue</i>
HA	Metodologia de Detecção de ROI - Haar-Like Features com Adaboost
LS	Metodologia de Detecção de ROI - LBP com SVM

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>14</b>
1.1 MOTIVAÇÃO	14
1.2 OBJETIVOS	16
1.2.1 Objetivo Geral	16
1.2.2 Objetivos Específicos	16
1.3 ORGANIZAÇÃO DA DISSERTAÇÃO	16
<b>2 FUNDAMENTAÇÃO TEÓRICA</b>	<b>18</b>
2.1 CONCEITOS DE PROCESSAMENTO DE IMAGENS	18
2.1.1 Reconhecimento Facial	19
2.1.2 <i>Local Binary Pattern</i>	20
2.1.3 <i>Haar-like Features</i>	21
2.1.3.1 Detecção de Faces com o Algoritmo de Viola e Jones	22
2.1.4 Morfologia Matemática Binária	23
2.1.4.1 Erosão	24
2.1.4.2 Dilatação	25
2.1.4.3 Abertura e Fechamento	26
2.1.5 Segmentação de Imagens	27
2.2 ALGORITMOS DE RECONHECIMENTO DE PADRÕES	29
2.2.1 <i>Support Vector Machines</i>	30
2.2.2 <i>Adaptive Boosting</i>	30
2.3 ALGORITMO DE BUSCA <i>VARIABLE NEIGHBORHOOD SEARCH</i>	31
2.4 BASE DE IMAGENS FACIAIS	33
2.4.1 Normas para Construção de Bases de Imagens	33
2.4.1.1 Poses e Posicionamento das Faces	34
2.4.1.2 Configuração do Ambiente	35
2.4.1.3 Configuração da Câmera e da Imagem	36
2.4.2 Exemplos de Bases de Imagens Faciais	37
2.5 DETECÇÃO DE REGIÕES DE INTERESSE	38
2.6 NORMALIZAÇÃO DE IMAGENS FACIAIS	40
2.6.1 Normas para Normalização de Imagens Faciais	41
2.7 TRABALHOS CORRELATOS	43
<b>3 DESENVOLVIMENTO</b>	<b>46</b>
3.1 CAPTURA DAS IMAGENS FACIAIS	46
3.1.1 Equipamentos	46
3.1.2 Posicionamento do Equipamento	48
3.1.3 Poses das Faces	50
3.1.4 Procedimento para Captura das Imagens	51
3.1.4.1 Codificação dos Arquivos de Imagem	52
3.2 DETECÇÃO DE OLHOS	53
3.2.1 <i>Haar-like Features</i> com Adaboost	54
3.2.2 LBP com SVM	54

3.3	NORMALIZAÇÃO FACIAL .....	56
3.3.1	Etapas da Normalização .....	57
3.3.1.1	Rotação .....	57
3.3.1.2	Redimensionamento .....	58
3.3.1.3	Posicionamento dos Olhos e Recorte da Imagem .....	58
3.3.1.4	Plano de Fundo .....	59
3.3.2	Normalização de Faces Laterais .....	60
3.3.3	Normalização Semiautomática .....	62
3.3.4	Normalização Automática .....	63
<b>4</b>	<b>EXPERIMENTOS E RESULTADOS .....</b>	<b>65</b>
4.1	BASE DE IMAGENS FACIAIS .....	65
4.2	DETECÇÃO DE OLHOS .....	70
4.3	NORMALIZAÇÃO FACIAL .....	73
4.3.1	Normalização Semiautomática Frontal .....	74
4.3.1.1	Normalização de Plano de Fundo .....	75
4.3.2	Normalização Semiautomática Lateral .....	76
4.3.3	Automática com Detecção de Olhos .....	78
<b>5</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>79</b>
5.1	DIFICULDADES ENCONTRADAS .....	79
5.2	CONCLUSÕES .....	79
5.3	TRABALHOS FUTUROS .....	80
5.4	PUBLICAÇÕES .....	81
	<b>REFERÊNCIAS .....</b>	<b>82</b>
	<b>Anexo A – TERMO DE ASSINATURAS PARA CAPTURA DAS IMAGENS FACIAIS</b>	
	<b>    PARA CONSTRUÇÃO DA BASE .....</b>	<b>86</b>
	<b>Anexo B – PUBLICAÇÕES .....</b>	<b>87</b>

# 1 INTRODUÇÃO

## 1.1 MOTIVAÇÃO

Nos últimos tempos, a visão computacional vem sendo utilizada com outras áreas de tecnologia, devido a sua capacidade de perceber o ambiente a sua volta. Com isso, podendo auxiliar a realização de diversas tarefas. O RF (reconhecimento facial), por exemplo, é uma área que utiliza a visão computacional como uma etapa em seu processo.

Com a evolução da Tecnologia da Informação, o reconhecimento de pessoas através de imagens tem se tornado comum. As imagens podem ser obtidas através de câmeras de monitoramento e, por questões de segurança, tem aumentado a demanda por sistemas automáticos de reconhecimento de faces. Existem muitos estudos e pesquisas que vêm sendo desenvolvidos pela comunidade científica com o objetivo de identificar pessoas usando somente a imagem da face. Para tanto, empregam-se técnicas e métodos científicos de diversas áreas, tais como visão computacional, reconhecimento de padrões, inteligência computacional e meta-heurísticas.

De modo geral, o reconhecimento de pessoas através de faces ainda se constitui uma tarefa desafiadora, pois as imagens que contém faces possuem alta variabilidade em tamanho, forma, cor e textura. Ainda mais, há a necessidade da identificação de faces invariante à posição, orientação ou variações de iluminação. Nas últimas décadas, inúmeras técnicas vêm sendo desenvolvidas para identificar faces. Desta forma, tem aumentada a demanda por métodos eficientes e robustos para a localização de faces humanas em imagens ou vídeos digitais, exigindo grande esforço computacional, técnicas inovadoras e algoritmos eficientes.

Por ser uma tarefa complexa, o RF é dividido em fases. Geralmente inicia-se pela captura de imagens. Esta etapa consiste em criar uma base com imagens faciais a serem utilizadas nas demais fases. A captura aparenta ser um processo simples, mas exige muito esforço. É recomendado um grande número de imagens por pessoa, em diversas condições e poses. Além disso, existem normas a serem seguidas para que as imagens sejam capturadas de forma a auxiliar o processo de RF.



Uma etapa que também visa auxiliar o processo de reconhecimento que é a normalização das faces. Esta fase tem como objetivo padronizar as faces utilizando técnicas de visão computacional. Para esta etapa também devem ser seguidas normas para padronização das imagens, que estão presentes nos mesmos documentos citados anteriormente.

Outra etapa importante do RF é a detecção de ROI (*Region of Interest* - Região de Interesse). As regiões podem ser os olhos, a face ou outras regiões que forneçam algum tipo de informação útil para o reconhecimento. A detecção pode ser utilizada como etapa inicial para a normalização, no caso deste trabalho a detecção dos olhos, onde os olhos servem de pontos de referência para esta tarefa.

Estas três fases (captura de imagens, detecção de ROIs e normalização facial) são objeto de pesquisa deste trabalho. Porém ainda existem outras etapas para o RF, como a extração de características, regiões onde existam características importantes para o reconhecimento. E com isso o reconhecimento propriamente dito pode ser executado.

O objetivo deste trabalho é construir uma base de imagens faciais com uma grande variabilidade de poses e iluminações, com o intuito de colaborar com pesquisas que necessitam de imagens faciais. Além da base de imagens ser composta pelas imagens capturadas, uma parte das imagens deve ser normalizada para complementar a base. Um nova metodologia de normalização automática é proposta neste trabalho. Para que isto funcione é preciso a detecção automática de ROI, então são propostas duas metodologias distintas para analisar qual seria mais eficiente em tempo e precisão.

Apesar de parecer um processo simples, existem as normas internacionais que regem os processos de captura e normalização de imagens faciais. No decorrer deste trabalho elas são explicadas de forma resumida e, também, como foram utilizadas e implementadas na realização do trabalho. Ainda seguindo as normas, foram feitas adaptações de modo a aumentar a diversidade da base. Além disto, é proposta uma nova metodologia de normalização lateral da imagens laterais, baseando-se nas normas existentes para normalização frontal.

Para a detecção de ROIs foram utilizadas técnicas de visão computacional, reconhecimento de padrões e de otimização de busca. Os filtros LBP (*Local Binary Pattern*) e Haar-Like Features são utilizados como pré-processamento para as imagens. Então elas são utilizadas para aprendizagem nos algoritmos SVM (*Support Vector Machines* - Máquinas de Vetores Suporte) e Adaboost (*Adaptive Boosting*). E para complementar as metodologias, é utilizado um algoritmo de otimização de busca, o VNS (*Variable Neighborhood Search*).

A execução deste trabalho resultou na formação de uma base de imagens faciais con-

tendo 3726 imagens, uma base normalizada com imagens frontais com 966 imagens e uma com imagens laterais com 276 imagens. A normalização das imagens frontais foi auxiliada pela detecção automática de olhos. Para a detecção foram comparadas duas técnicas, onde em uma delas o melhor caso foi de 99,2% de precisão e na segunda, na média, 94,9%.

## 1.2 OBJETIVOS

### 1.2.1 OBJETIVO GERAL

O objetivo principal deste trabalho é desenvolver uma metodologia para a normalização automática de imagens faciais baseada na detecção automática de olhos. Para isto, foi construída uma base de imagens a ser normalizada.

### 1.2.2 OBJETIVOS ESPECÍFICOS

Como objetivos específicos tem-se:

1. Propor uma nova metodologia para captura de imagens faciais baseada em normas internacionais e com algumas adaptações em relação as normas.
2. Construir uma base de imagens faciais com várias posições, orientações e sob algumas condições de iluminação padronizada;
3. Comparar duas metodologias para detecção automática de olhos utilizando técnicas de visão computacional, reconhecimento de padrões, algoritmos de busca local e otimização;
4. Desenvolver um algoritmo para normalização de imagens faciais, baseada em normas internacionais.

## 1.3 ORGANIZAÇÃO DA DISSERTAÇÃO

O Capítulo 2 desta dissertação discorre sobre assuntos que fundamentam o trabalho. Breves explicações e exemplos das técnicas utilizadas são apresentadas. As normas de captura e de normalização são descritas. Também são mostrados alguns trabalhos correlatos com a dissertação.

No Capítulo 3 apresenta-se a metodologia desenvolvida neste trabalho, especificamente dois algoritmos de detecção de ROI com a ideia central semelhante. E também é explicado como as normas de captura e normalização foram seguidas para a execução deste trabalho.

No próximo capítulo, Capítulo 4, são apresentados experimentos e resultados sobre os três principais assuntos deste trabalho: captura de imagens, comparação entre as duas técnicas de detecção de ROIs e normalização de imagens faciais.

Por fim as considerações finais, conclusões e trabalhos futuros são apresentados no Capítulo 5. Neste Capítulo também são apresentados artigos publicados durante o desenvolvimento deste trabalho.

## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados os principais conceitos sobre técnicas utilizadas neste trabalho. Inicialmente uma visão geral sobre algumas técnicas de processamento de imagens, na Seção 2.1. Algoritmos de reconhecimento de padrões e de busca também foram utilizadas na metodologia deste trabalho e estão explicadas nas seções 2.2 e 2.3, respectivamente.

Nas seções 2.4 e 2.6 são resumidas as principais normas referentes a captura e normalização de imagens, que foram utilizadas na metodologia deste trabalho, apresentadas nos documentos *Police Standard for Still Digital Image Capture and Data Interchange of Facial/Mugshot and Scar, Mark and Tattoo Images* (ISLAM, 2007), *Best Practice Recommendation for the Capture of Mugshots - NIST Best Practices* (NIST, 2007), *Capture and Equipment Assessment for Face Recognition Systems* (FACIAL IDENTIFICATION SCIENTIFIC WORKING GROUP, 2011) e no relatório técnico *Face Recognition Format for Data Interchange* (GRIFFIN, 2003).

Trabalhos correlacionados são apresentados na ultima seção deste capítulo (Seção 2.7). Estes apresentam bases de imagens existentes na literatura, sem normalização e outras normalizadas. Trabalhos de detecção de regiões em faces também são apresentadas, mostrando a metodologia utilizada pelos autores e também os resultados obtidos.

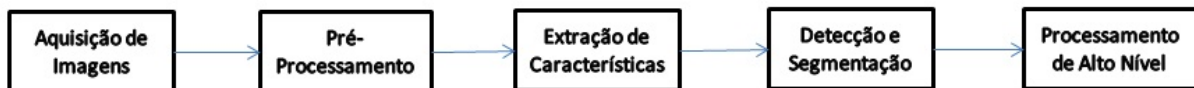
### 2.1 CONCEITOS DE PROCESSAMENTO DE IMAGENS

Assim como os seres humanos percebem o mundo ao seu redor através da visão, as máquinas também podem tirar informações do ambiente a sua volta e simular a percepção humana. A área responsável por isto é a visão computacional. Isto pode ser feito através de sensores como, por exemplo, máquinas fotográficas, raio-X, ressonância magnética, entre outros. Com isso, os computadores podem aprender, reconhecer cenários e objetos, explorar salas e até mesmo interagir com humanos.

A visão computacional pode ser aplicada em diversas áreas, como exemplo: na área industrial, para orientação de robôs em uma linha de produção; a medicina pode utilizar a visão

computacional para analisar imagens capturadas por radiografia e tomografia; para segurança pública; entre outros. Neste trabalho, a área em questão é a de segurança, tratada através do reconhecimento facial.

Grande parte dos sistemas de visão computacional possuem uma série de tarefas que seguem uma sequência, como mostrado na Figura 1. A aquisição de imagens é a primeira tarefa para qualquer sistema de visão computacional, através de sensores as imagens são capturadas para a criação de uma base de imagens para posterior processamento. Após isso, no pré-processamento, as imagens são preparadas através da aplicação de filtros, redução de ruídos, normalização, entre outros, com o objetivo de melhorar a imagem de acordo com as necessidades da sua aplicação, auxiliando na extração de características. Este processo é responsável pela aquisição de informações contidas nas imagens, que podem ser utilizadas na classificação de imagens, por exemplo. Na etapa de detecção e segmentação são identificadas ROIs, onde o objeto a ser estudado está presente. E, por fim, o processamento em alto nível, que é responsável pela classificação dos dados retirados das imagens.



**Figura 1: Etapas de um Sistema de Visão Computacional.**

**Fonte: Autoria Própria.**

As principais etapas utilizadas neste trabalho são captura de imagens, pré-processamento (normalização) e detecção. Estes estão explicados nas próximas seções, onde são apresentados conceitos e as principais técnicas já utilizadas em trabalhos relacionados, principalmente com o RF.

### 2.1.1 RECONHECIMENTO FACIAL

O RF é um ramo da visão computacional que vem crescendo e está sendo aplicado em diversas áreas. Este se tornou uma importante aplicação na área de segurança. Na área de segurança pública, o RF é utilizado para identificar suspeitos que praticaram algum tipo de ato ilícito, em especial aqueles que já têm ficha policial cuja face se encontra em uma base de imagens da polícia. Além disto, o RF também é utilizado em empresas para controle de acesso a lugares restritos (REISFELD; YESHURUN, 1998; ZHAO et al., 2003).

O RF abrange diversas etapas, como mostrado na Figura 1 (CHANG et al., 2003). Estas etapas vão desde a captura de imagens para treinamento do sistema até a fase de reconhe-

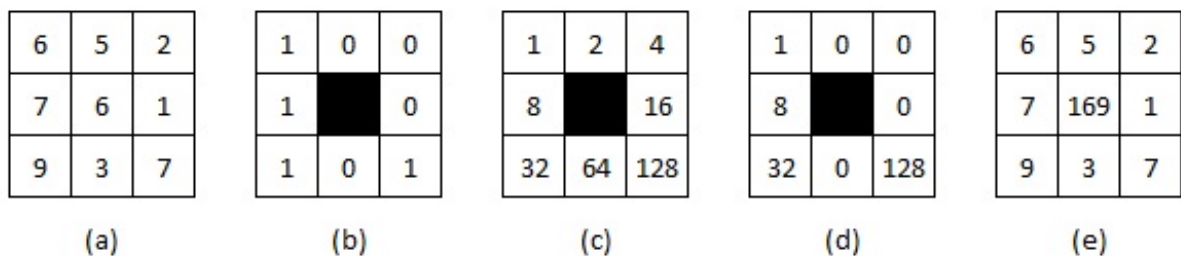
cimento em si, incluindo detecção de ROI, normalização da face, extração de características e a comparação com as imagens armazenadas em uma base, ou seja, o próprio reconhecimento (REISFELD; YESHURUN, 1998; ZHAO et al., 2003). A etapa de detecção consiste em encontrar a localização espacial do que está sendo procurado. Com isto, é possível realizar a etapa de normalização, que é responsável por padronizar a posição e tamanho de elementos da face em relação à imagem. Em seguida, pode ser feita a extração de características para a identificação, comparação ou reconhecimento das imagens dentro de uma base.

Dentre estas etapas, a mais importante é a detecção dos olhos, pois estes são o elemento facial mais relevante para o reconhecimento. O posicionamento dos olhos é utilizado em outras etapas. Na extração de características, por exemplo, podem conter informações que diferenciem pessoas (ZHAO et al., 2003). A detecção dos olhos é o primeiro passo para a normalização, esta etapa é discutida na Seção 2.6.

De modo geral, este trabalho pode ser classificado como um pré-processamento para o RF. As etapas desenvolvidas são as iniciais, captura de imagens e normalização facial. Para a normalização é necessária a detecção de ROI, no caso deste trabalho os olhos. Através destas etapas, as imagens estão melhor preparadas para passar pelo processo de reconhecimento.

### 2.1.2 LOCAL BINARY PATTERN

LBP é um filtro de textura proposto por Ojala et al. (1996). Este filtro foi baseado no modelo de análise de texturas de Wang e He (apud OJALA et al., 1996), onde a região de textura era representada por 8 elementos (uma região de vizinhança de 3x3 pixels). O LBP é uma versão onde cada elemento pode receber 2 valores (0 e 1), com isso, o número de possíveis texturas é 256 ( $2^8$ ) (OJALA et al., 2002). O processo de aplicação do filtro LBP é mostrado na Figura 2.



**Figura 2: Procedimento do LBP.**

**Fonte: Ojala et al. (1996).**

O procedimento da Figura 2 é repetido para cada pixel da imagem, como explicado

nas etapas a seguir.

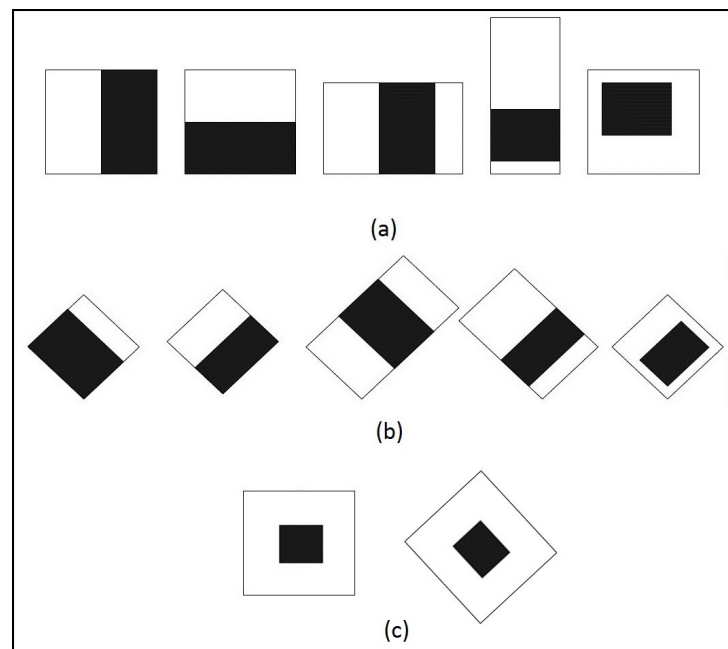
1. Inicialmente uma região da imagem (neste exemplo uma região 3x3) é analisada a partir do pixel central, valor 6 na Figura 2 (a);
2. Cada valor da região é analisado, recebendo valor 1 caso seja maior ou igual ao pixel central e 0 caso contrário, imagem 2 (b), resultando numa matriz binária;
3. Uma matriz auxiliar, do mesmo tamanho da região, é criada, sendo que seus valores seguem a potência de 2 (2 elevado à posição da matriz, circundando o valor central), como visto na Figura 2 (c). O valor central desta matriz não será utilizado;
4. Cada valor da matriz binária é multiplicado pelos valores da matriz auxiliar em suas respectivas posições, resultando na matriz da Figura 2d;
5. Por fim, os valores são somados e atribuídos ao pixel central da região da imagem, como observado na imagem 2 (e).

O LBP é geralmente utilizado como um filtro de pré-processamento para reconhecimento de padrões. Para alguns casos, quando utilizado isoladamente o LBP não é eficiente, por isso é utilizado em conjunto com algum outro algoritmo. O que mais se encontra na literatura é a utilização do LBP com o SVM (Seção 2.2). Porém, também pode ser utilizado sem ajuda de outro algoritmo e mesmo assim se tornar eficiente, como no caso do trabalho de Polli et al. (2011) onde o LBP foi utilizado para extração de dados biométricos na identificação da palma da mão.

### 2.1.3 HAAR-LIKE FEATURES

*Haar-like Features* foi proposto por Viola e Jones (2001) e consiste em uma modificação do *Haar Wavelets* proposto por Papageorgiou (PAPAGEORGIOU et al., 1998 apud VIOLA; JONES, 2001) para detecção de objetos. É uma forma de avaliar as imagens usando características ao invés dos valores dos pixels diretamente, ou seja, meta-características que não estão explícitas nas intensidades dos pixels da imagem.

O *Haar-like Features* utiliza regiões conexas retangulares (Figura 3), que se deslocam pelos pixels da imagem realizando o cálculo das características. Primeiramente calcula-se a soma ponderada dos pixels da região branca, então, este valor é subtraído da soma ponderada dos pixels da região preta, onde cada posição recebe como peso o seu valor de intensidade (VIOLA; JONES, 2001).



**Figura 3: Modelos de *Haar-like Features*.**

**Fonte: Modificado da Documentação do OpenCV <sup>1</sup>.**

Na Figura 3, é possível notar várias formas retangulares que podem ser utilizadas para calcular o *Haar-like Features*, utilizadas para obter características das bordas (Figura 3 (a)), linhas (Figura 3 (b)) e regiões centrais (Figura 3 (c)). Mais de um modelo pode ser utilizado em conjunto para realizar a extração das características.

A principal vantagem deste recurso é a velocidade computacional de cálculo. Porém, assim como o LBP, o *Haar-like Features* se torna mais eficiente quando utilizado com algum algoritmo de reconhecimento de padrões. Viola e Jones (2001) utilizaram estas características junto com o Adaboost (Seção 2.2) para desenvolver um algoritmo de detecção de faces mostrado a seguir.

### 2.1.3.1 DETECÇÃO DE FACES COM O ALGORITMO DE VIOLA E JONES

Esta abordagem para detecção de faces foi proposta por Viola e Jones (VIOLA; JONES, 2001). Suas principais contribuições marcantes são: a filtragem das imagens através do filtro *Haar-like Features*, que faz seu cálculo de forma rápida; a construção de um classificador selecionando um conjunto pequeno de características através do Adaboost; e, por último, a construção de um classificador complexo numa estrutura cascata que aumenta drasticamente a velocidade do detector focando apenas nas regiões promissoras.

Este algoritmo realiza a extração de características percorrendo a região de busca,



como mostra a Figura 4. As regiões com retângulos pretos representam o *Haar-like Features* realizando o cálculo das características. Cada uma destas regiões, já com suas características em um vetor, são enviadas para classificação com o algoritmo Adaboost (Seção 2.2). Com isso, é possível percorrer a imagem classificando as regiões até que o Adaboost indique sucesso na sua classificação, ou seja, indicando que determinada região é uma face.



**Figura 4: Processo de extração de características do *Haar-like*.**

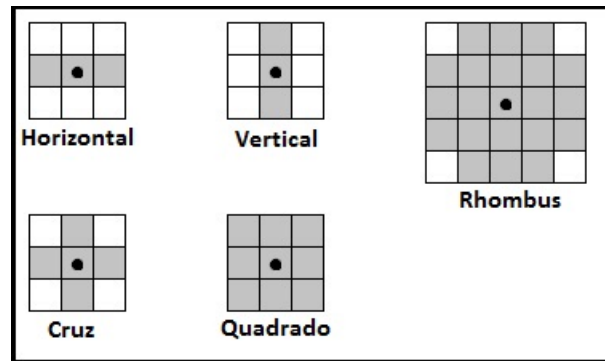
**Fonte: Retirado do site Makematics <sup>2</sup>**

Uma metodologia baseada no algoritmo de Viola e Jones foi desenvolvida neste trabalho com algumas adaptações para detectar outras ROIs, diferentes da face.

#### 2.1.4 MORFOLOGIA MATEMÁTICA BINÁRIA

A Morfologia (do grego *morphê*-forma e *logos*-ciência) trata do estudo das formas que a matéria pode tomar e geralmente é usada para trabalhar com a forma e estrutura de plantas e animais (FACON, 1996). Da mesma maneira, a Morfologia Matemática funciona como uma ferramenta para extração de componentes que são utilizadas para representar formas de regiões de imagens como, por exemplo, bordas e esqueletos.

O Elemento Estruturante é o principal componente da Morfologia Matemática, podendo ser de vários formatos, como mostrado na Figura 5. Consiste em um conjunto definido, que vai interagir com cada região da imagem modificando sua morfologia e, com isto, tentar obter uma conclusão sobre cada uma delas. Um desafio da Morfologia Matemática está em escolher qual o melhor Elemento Estruturante para determinado problema.



**Figura 5: Exemplos de Elementos Estruturantes.**

**Fonte: Autoria Própria.**

Na Figura 5 são mostrados 5 exemplos diferentes de Elementos Estruturantes: Horizontal, Vertical, Cruz, Quadrado e Rhombus. Porém, podem ser criadas várias outras formas de acordo com a necessidade da aplicação. Em cada um dos elementos da figura, nota-se um ponto preto representando o ponto que será modificado pelas transformações, o qual não precisa estar necessariamente no centro, podendo gerar outras variações.

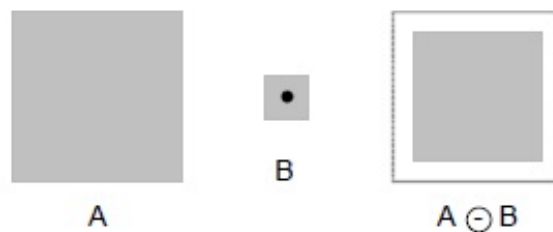
Além da Elemento Estruturante, existem dois operadores importantes que formam os pilares da Morfologia Matemática, a Erosão e a Dilatação. Embora a Morfologia Matemática seja caracterizada pela filtragem não-linear, estes operadores não são consideradas filtros, mas são operações elementares para filtros mais complexos e poderosos (FACON, 1996). Como exemplo destes filtros pode-se citar a Abertura e o Fechamento, explicados no decorrer desta seção. Os operadores de Erosão e Dilatação são apresentados a seguir.

#### 2.1.4.1 EROSÃO

Para imagens binárias, a Erosão ( $\ominus$ ) de uma imagem  $A$  por um Elemento Estruturante  $B$  é definida na Equação 1:

$$A \ominus B = \{x \in A : B_x \subset X\} \quad (1)$$

Em resumo, a Equação 1 o Elemento Estruturante  $B$ , transladado de  $x$ , deve percorrer toda a imagem  $A$  posicionando o Ponto Central no pixel  $x$ , obtendo-se como resultado o conjunto de pontos  $x$  contido em  $A$  (MARQUES FILHO; VIEIRA NETO, 1999). As imagens da Figura 6 facilitam a compreensão do funcionamento da operação de Erosão. É possível notar a ação de Erosão do Elemento Estruturante Quadrado (Figura 5) sobre a imagem  $A$ . O Elemento Estruturante varre a imagem efetuando a comparação da região atual com o elemento estruturante de acordo com a Equação 1. Todos os pixel do quadrado (Figura 6) em que a equação é verdadeira ele afetado pelo elemento estruturante. Ou seja, o tamanho do quadrado é reduzido em todos os lados na quantidade de pixels que o elemento estruturante possui.



**Figura 6: Exemplo do Processo de Erosão.**

**Fonte: (MARQUES FILHO; VIEIRA NETO, 1999).**

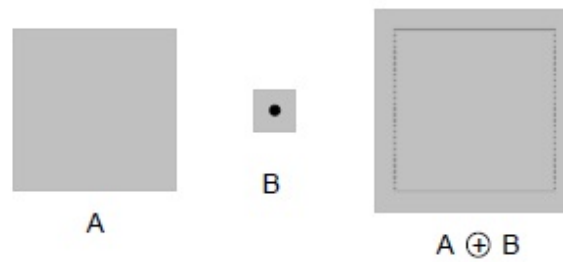
Os principais efeitos da Erosão sobre as imagens são: diminuir partículas ou ruídos, até mesmo eliminá-los caso sejam menores do que o Elemento Estruturante; aumentar regiões que representem buracos; e separar algumas partículas próximas. (FACON, 1996)

#### 2.1.4.2 DILATAÇÃO

A Dilatação ( $\oplus$ ) de uma imagem binária  $A$  por um Elemento Estruturante  $B$  é representada como mostrado na Equação 2.

$$A \oplus B = \{x \in A : B_x \cap X \neq 0\} \quad (2)$$

A Dilatação pode ser dita como o dual da Erosão. Ou seja, a Dilatação consiste em deslocar a reflexão do Elemento Estruturante  $B$  em  $x$ , obtendo-se como resultado o conjunto de todos os deslocamentos onde a interseção de  $B_x$  e  $A$  inclui ao menos um elemento diferente de 0 (MARQUES FILHO; VIEIRA NETO, 1999). Este processo pode ser melhor compreendido observando-se as imagens da Figura 7. Neste caso, o quadrado vai aumentar de tamanho em todos os lados, de acordo com o tamanho do elemento estruturante.



**Figura 7: Exemplo do Processo de Dilatação.**

**Fonte: (MARQUES FILHO; VIEIRA NETO, 1999).**

Como a Dilatação é o dual da Erosão, os efeitos da aplicação desta operação na imagem são contrários: aumentar partículas, preencher elementos, como buracos e conectar regiões próximas (FACON, 1996).

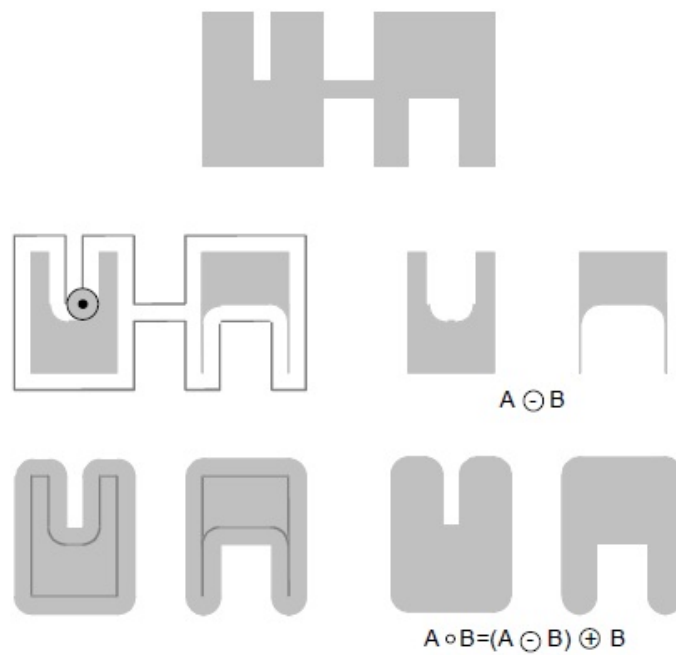
#### 2.1.4.3 ABERTURA E FECHAMENTO

O processo de abertura é feito através da Erosão de  $A$  por  $B$  seguida pela Dilatação do resultado por  $B$  (GONZALEZ; WOODS, 2006). A Abertura pode ser utilizada para separar partículas ou até eliminá-las, mantendo outras regiões semelhantes às originais. A imagem resultante é considerada mais regular que a original, porém com menos detalhes (FACON, 1996).

Na Figura 8 é exemplificado o processo de Abertura. Neste exemplo, é possível notar que a imagem original passa pelo processo de Erosão por um Elemento Estruturante semelhante ao Rhombus (Figura 5) e em seguida o resultado é Dilatado pelo mesmo elemento, resultando na Abertura.

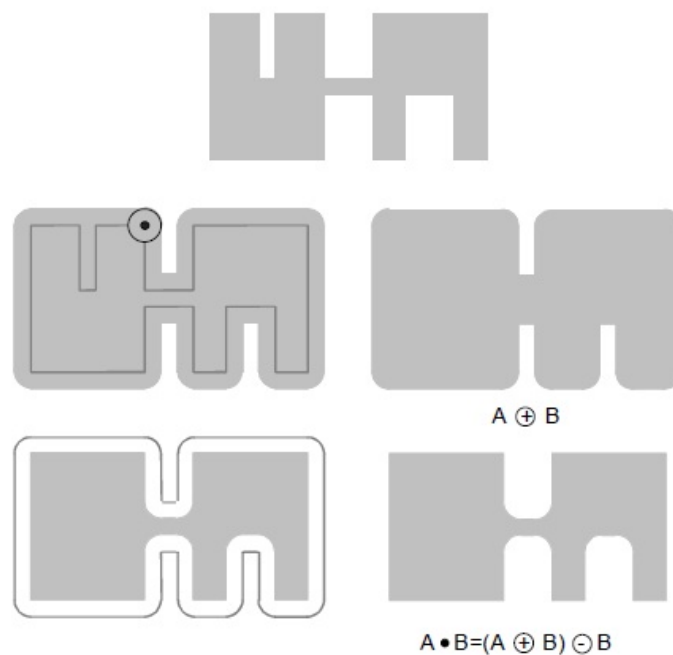
O Fechamento é a operação dual da Abertura, ou seja, primeiramente é realizada a Dilatação de  $A$  por  $B$  e em seguida a Erosão pelo Elemento Estruturante  $B$  (GONZALEZ; WOODS, 2006). O Fechamento pode ser utilizado para suavizar fronteiras, preencher buracos ou emendar partículas da imagem, sem muita alteração dos elementos que são maiores que o Elemento Estruturante. A imagem resultante, assim como para a abertura, é mais regular e com tipos de detalhes diferentes (FACON, 1996).

A Figura 9 exemplifica o processo de Fechamento, representando a Dilatação seguida da Erosão da imagem pelo Elemento Estruturante Rhombus, resultando na imagem Fechada.



**Figura 8: Exemplo do Processo de Abertura.**

**Fonte: (MARQUES FILHO; VIEIRA NETO, 1999).**



**Figura 9: Exemplo do Processo de Fechamento.**

**Fonte: (MARQUES FILHO; VIEIRA NETO, 1999).**

### 2.1.5 SEGMENTAÇÃO DE IMAGENS

A segmentação de imagens tem como objetivo separar uma imagem em regiões distintas por meio de algum critério (GONZALEZ; WOODS, 2006). Este processo simplifica a

imagem para facilitar sua análise. A segmentação pode ter como resultado final conjuntos de regiões ou de bordas, de acordo com a similaridade dos pixels. Por exemplo, a Figura 10 mostra um processo de segmentação, sendo que a primeira imagem é a original, a segunda está segmentada em três áreas e, por fim, a terceira possui somente uma ROI.



**Figura 10: Exemplo de Segmentação.**

**Fonte: Autoria Própria.**

Os critérios utilizados para segmentar uma imagem podem ser, por exemplo: limiares de cor ou intensidade, conectividade entre pixels, características de vizinhança, bordas, histogramas, entre outros (GONZALEZ; WOODS, 2006).

Segmentar imagens por limiares de cor é a forma mais comum na literatura. Limiares de cor são valores definidos onde a imagem é segmentada em regiões através da comparação dos valores de todos os pixels com estes limiares. Assim, a imagem pode ser dividida em várias regiões ou em apenas duas regiões, caso tenha somente um limiar.

A segmentação através da cor pode ser feita no espaço de cores HSI (*Hue-Saturation-Intensity* - Matiz-Saturação-Intensidade). O componente matiz carrega a maior parte de informação de cores, por isso é importante para a segmentação. E a componente saturação é utilizada para isolar algumas regiões na imagem em conjunto com a componente matiz. Por outro lado, a intensidade não é utilizada na segmentação de imagens, pelo fato de não carregar informação de cor (GONZALEZ; WOODS, 2006).

*Region Growing* é o nome dado para a principal forma de segmentação de imagens baseada em conectividade entre pixels (GONZALEZ; WOODS, 2006). Este algoritmo tem como objetivo agrupar pixels ou regiões de acordo com critérios pré-definidos. O algoritmo *Region Growing* é executado estabelecendo alguns pontos iniciais na imagem e então cada ponto é comparado com os pixels da vizinhança. Através da similaridade entre os pixels, os pontos iniciais vão agregando mais pixels e formando regiões maiores, até que toda a imagem

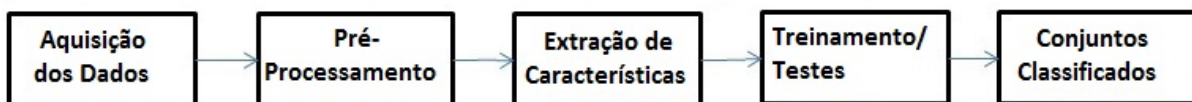
seja explorada e segmentada.

## 2.2 ALGORITMOS DE RECONHECIMENTO DE PADRÕES

As técnicas de reconhecimento de padrões utilizam a indução como princípio básico, ou seja, realizam classificações genéricas referentes a determinado conjunto de informações. Inicialmente, o algoritmo precisa aprender o que vai classificar, esta etapa é chamada de treinamento (aprendizagem). O treinamento é feito apresentando vários exemplos do que se deseja classificar, no decorrer de algumas iterações as variáveis do algoritmo são atualizadas de acordo com o que é apresentado, de forma a se adaptar ao conjunto de informações que está sendo apresentado para classificá-las de forma correta. Um tipo de treinamento é o supervisionado, onde, além de apresentar o conjunto de treino, também é informado ao algoritmo a que classe determinado conjunto pertence. Outro tipo é o não-supervisionado, neste caso o algoritmo tem que aprender de acordo com os padrões apresentados, sem saber a que classe cada um pertence.

Com o algoritmo treinado é possível realizar a validação ou teste do algoritmo, onde outro conjunto de dados é apresentado para classificação e então é possível verificar se as respostas estão corretas. Caso contrário, pode ser treinado novamente com modificações de parâmetros, por exemplo. Após uma validação bem sucedida, o algoritmo está pronto para ser utilizado como classificador em alguma aplicação.

As etapas básicas de um algoritmo de reconhecimento de padrões podem ser vistas na Figura 11. A primeira consiste na obtenção de vários dados de exemplo para o algoritmo. Em seguida, eles passam pelo Pré-Processamento, onde são refinados e, então, as características relevantes para a classificação são extraídas. Na etapa seguinte, ocorre a execução do algoritmo em si, tanto para treinamento quanto para testes. E, com isso, os dados estão classificados.



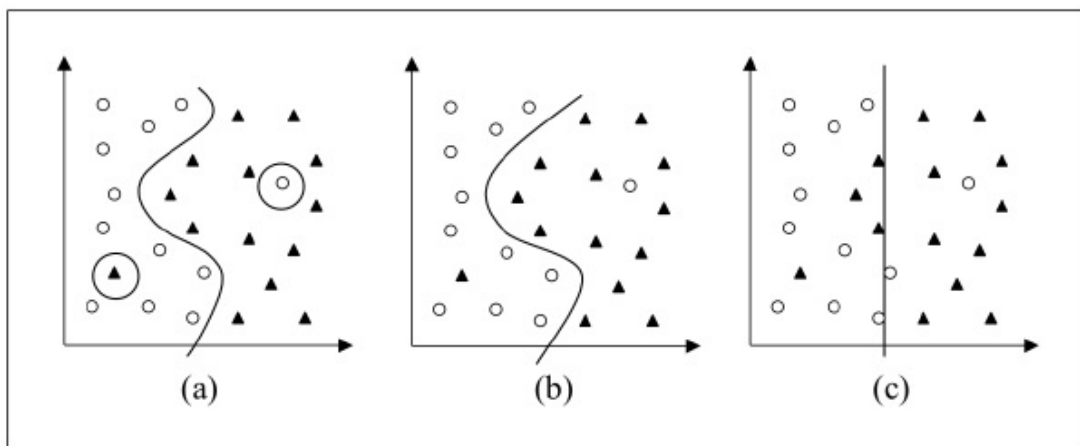
**Figura 11: Etapas de um algoritmo de Reconhecimento de Padrões.**

**Fonte: Autoria Própria.**

Existem vários algoritmos de reconhecimento de padrões, sendo que os apresentados nesta seção são as Máquinas de Vetores de Suporte e o Adaboost, utilizados no trabalho.

### 2.2.1 SUPPORT VECTOR MACHINES

As SVM (VAPNIK, 1995) são caracterizadas pelo aprendizado estatístico, isto é, o algoritmo estabelece relações matemáticas para a criação de um classificador de acordo com o conjunto de dados que lhe foi apresentado para treinamento. O SVM é utilizado para resolver problemas binários, seu objetivo é encontrar uma função, através do treinamento, que divida as duas classes a partir dos dados apresentados. Esta função gera um hiperplano que é responsável por separar as classes, um exemplo de classes separadas por um hiperplano é apresentado na Figura 12.



**Figura 12: Exemplo de classes separadas por um hiperplano.**

**Fonte: (LORENA; CARVALHO, 2007)**

Neste exemplo (Figura 12) existem 3 hiperplanos diferentes formados pelo treinamento de um SVM. A Figura 12 (a) e (b) mostra dois hiperplanos diferentes com apenas 2 ruídos na classificação, ou seja, um dado para cada conjunto foi classificado de forma incorreta. Já o hiperplano da Figura 12 (c) ignorou os dados de classes distintas que se encontram muito perto, neste caso o treinamento pode ser mais rápido, porém ocorrem muitos erros. Um hiperplano ótimo é quando todos os dados são classificados corretamente.

### 2.2.2 ADAPTATIVE BOOSTING

Adaboost, simplificado de *Adaptative Boosting* (Estímulo Adaptativo), proposto por Freund e Scharpire (1997), é um meta-algoritmo de classificação. Recebe este nome pois é adaptativo no sentido de que a classificações subsequentes são adaptadas de acordo com as anteriores. O algoritmo Adaboost é sensível a ruídos nos dados e a existência de dados discrepantes.



A Figura 13 mostra um pseudo-código do Adaboost. O algoritmo é executado por algumas iterações, mantendo uma distribuição de probabilidade  $D$  e a cada iteração o Adaboost é treinado com relação a esta distribuição. O resultado do treinamento gera a hipótese  $h$ , onde  $Y$  são as classes e  $X$  as características de entrada. Após cada iteração a taxa de erro  $\varepsilon$  é calculada e a distribuição  $D$  aumentada de acordo com classificações erradas. Depois de todas as rodadas a hipótese final é escolhida por ponderação (SCHWENK; BENGIO, 2000).

**Inicialização:**

- $B = \{(i, y) : i \in \{1, \dots, N\}, y \neq y_i\}$
- $D_1(i, y) = 1/|B|$  para todo  $(i, y) \in B$

**Repita:**

- (1) Treinar a rede respeitando a distribuição  $D_t$  e obter hipótese  $h_t$ :  
 $X \times Y \rightarrow [0, 1]$
- (2) Calcular a perda de  $h_t$ :  
 $\varepsilon_t = \frac{1}{2} \sum_{(i,y) \in B} D_t(i,y) (1 - h_t(x_i, y_i) + h_t(x_i, y))$
- (3)  $\beta_t = \varepsilon_t / (1 - \varepsilon_t)$
- (4) atualizar distribuição  $D_t$ :  

$$D_{t+1}(i, y) = (D_t(i, y) / Z_t) \beta_t^{1/2(1 + h_t(x_i, y_i) - h_t(x_i, y))}$$
 onde  $Z_t$  é uma constante normal

**Hipótese final:**

$$f(x) = \operatorname{argmax}_{(y \in Y)} \sum_t (\log 1/\beta_t) h_t(x, y)$$

Figura 13: Pseudo-código do *Adaptive Boosting*.

Fonte: A autoria Própria.

### 2.3 ALGORITMO DE BUSCA *VARIABLE NEIGHBORHOOD SEARCH*

Algoritmos de busca são responsáveis por encontrar a melhor solução possível dentro de um espaço de busca. Geralmente, a melhor solução é encontrada depois de diversas tentativas buscando a melhor solução.

Dentre eles, o mais simples é o de busca exaustiva, onde todas as possíveis soluções são analisadas, demandando muito tempo de processamento. Outros algoritmos são otimizados para que não realizem uma busca exaustiva. Estes procuram encontrar soluções de boa qualidade para o problema. Esta qualidade da solução é conhecida como *fitness*.

Um exemplo de algoritmo de busca é a heurística VNS, proposto por Mladenovic e Hansen (1997). Este tem como princípio básico pesquisar no espaço de busca utilizando o princípio da vizinhança. O VNS prossegue a busca através de método descida, para um local mínimo e, então, explora uma série de vizinhanças pré-determinadas. A partir destes locais

mínimos, tenta escapar dos vales com a perturbação. Ao contrário de outras Metaheurísticas (*Particle Swarm Optimization* (EBERHART; KENNEDY, 1995) ou *Ant Colony Optimization* (DORIGO; COLORNI, 1996), por exemplo), não segue nenhuma trajetória, mas sim, explora cada vez mais as vizinhanças distantes em relação ao solução atual. Muda para uma nova vizinhança somente se encontrar uma melhoria em relação a atual. Além disso, numa determinada vizinhança, a rotina da busca local é repetidamente aplicada para encontrar uma ótima local.

A estrutura de vizinhança pode ser definida como um conjunto finito de pré-selecionado de estruturas de vizinhanças, representada por  $N_k$  onde  $k = 1, \dots, k_{max}$  e o conjunto de soluções de  $k$ -ésimo vizinhança de  $x$ , por  $N_k(x)$ . O pseudocódigo básico do VNS é apresentado na Figura 14 (HANSEN; MLADENOVIC, 2001):

*Inicialização:*

- Selecione um conjunto de estruturas de vizinhança  $N_k$ ;
- Encontre uma solução inicial  $x$ ;
- *Repita* até atender critério de parada:
  - (1)  $k = 1$ ;
  - (2) *Repita* até  $k = k_{max}$ :
    - 1 - Perturbação de variáveis;  
(Gera um ponto  $x'$  aleatoriamente da  $k$ -ésima vizinhança de  $x$  onde  $x'$  pertence  $N_k(x)$ );
    - 2 - Busca local usando como a solução inicial  $x'$  selecionando  $x''$  como ótimo local;
    - 3 - Mudança de vizinhança:  
Se ( $x''$  melhor que  $x$ ) então  
mudar para nova vizinhança ( $x = x''$ )  
continua busca  $N_i(k - 1)$
    - Senão  
 $k = k + 1$

**Figura 14: Pseudo-código do *Variable Neighborhood Search*.**

**Fonte: Autoria Própria.**

O critério de parada pode ser determinado por um número máximo de iterações ou número máximo de iterações entre duas melhorias (HANSEN; MLADENOVIC, 2001). A vantagem do VNS é que este permite encontrar soluções ótimas ou próxima à ótima no tempo moderado.

## 2.4 BASE DE IMAGENS FACIAIS

Visão computacional descreve sistemas artificiais que obtêm informações de imagens. Portanto, a construção de uma base de imagens é um passo importante para a área de visão computacional. Existem diversas sub-áreas que necessitam de uma base para treinamento de algoritmos e também para estudos das imagens. Por se tratar de um assunto relacionado ao RF, a base de imagens deste trabalho deve conter faces de pessoas.

Para o RF são utilizadas diversas imagens, por isto uma base de imagens deve conter faces de diversas pessoas e em diversas condições. Isto é necessário em casos de segurança, por exemplo, quando um suspeito é fotografado ele raramente vai estar olhando para a câmera e a iluminação pode não contribuir. Para isso algumas variações de imagens são necessários, como exemplo a pose (frontal, lateral, inclinada), a iluminação, a oclusão, entre outras (REISFELD; YESHURUN, 1998).

A captura de imagens para a construção de uma base aparenta ser um processo simples, mas existe uma sequência de passos que devem ser seguidos. Dentre estes passos esta a organização do ambiente, o que inclui o posicionamento da iluminação, da câmera e da própria pessoa. Outro fator que dificulta o processo de captura é posicionar a face da pessoa de acordo as variações desejadas. Para padronizar os passos da captura de imagens faciais existem algumas normas internacionais, apresentadas a seguir.

### 2.4.1 NORMAS PARA CONSTRUÇÃO DE BASES DE IMAGENS

Existem vários documentos que estabelecem normas para a captura de imagens. Os principais são: *Police Standard for Still Digital Image Capture and Data Interchange of Facial/Mugshot and Scar, Mark and Tattoo Images* (ISLAM, 2007), *Best Practice Recommendation for the Capture of Mugshots - NIST Best Practices* (NIST, 2007), *Capture and Equipment Assessment for Face Recognition Systems* (FACIAL IDENTIFICATION SCIENTIFIC WORKING GROUP, 2011) e o relatório técnico *Face Recognition Format for Data Interchange* (GRIFFIN, 2003).

O documento de Islam (2007) apresenta diversas normas referentes à captura de imagens faciais, imagens faciais com cicatrizes ou marcas, imagens com tatuagens e também de corpo inteiro. Porém, as normas descritas a seguir são referentes somente a imagens de face, foco deste trabalho, sem cicatrizes, marcas e tatuagens. Em grande parte das normas de seu documento, Islam (2007) baseou-se no documento da NIST (*National Institute os Standarts and Technology - Instituto Nacional de Padrões e Tecnologia*).

O documento do NIST (NIST, 2007), assim como os de *Facial Identification Scientific Working Group* (FACIAL IDENTIFICATION SCIENTIFIC WORKING GROUP, 2011) e Griffin (2003), tratam somente sobre a captura de faces. Dentre eles, o mais conhecido e utilizado é o da NIST. Os três documentos descrevem vários detalhes sobre a captura como, por exemplo: poses das pessoas, acessórios, formas de iluminação do ambiente de captura, cor do plano de fundo das imagens, configurações gerais da câmera e formato de compressão do arquivo, entre outros. O documento de Griffin (2003) é o mais simplificado em relação a captura de imagens, mas a normalização das imagens é tratada com mais precisão. Este assunto é tratado na Seção 2.6.

Existem algumas pequenas variações entre os documentos citados, mas no geral eles tratam as normas de forma muito semelhante. Um compilado geral destes quatro documentos citados é descrito no decorrer desta Seção.

#### 2.4.1.1 POSES E POSICIONAMENTO DAS FACES

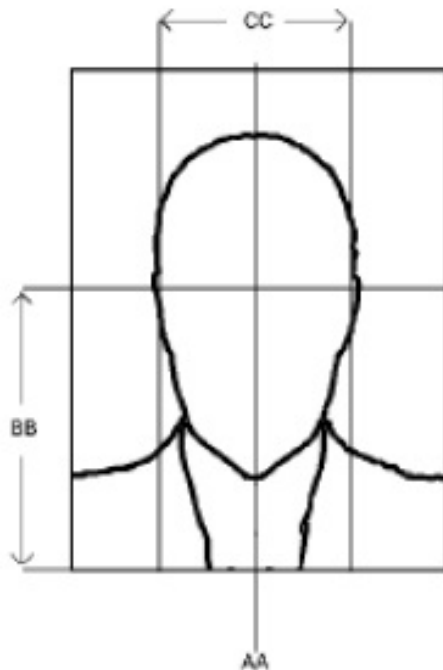
Em relação à pose, um requisito mínimo é a aquisição imagens frontais da face, em especial para bases que serão utilizadas por sistemas automatizados. Neste contexto, a cabeça inteira deve ser apresentada, inclusive o cabelo, pescoço, ombros e orelhas. O indivíduo deve olhar diretamente para a câmera e as íris e pupilas devem aparecer claramente. Caso o indivíduo use óculos, ao menos uma foto deve ser capturada sem eles. Nas imagens frontais, a inclinação máxima permitida pelas normas é de +/- 5 graus em qualquer direção (ISLAM, 2007; GRIFFIN, 2003). Nada deve interferir na face como, por exemplo, cabelos compridos devem ser amarrados ou colocados para trás da orelha.

Islam (2007) menciona algumas poses complementares que podem ser capturadas, tal como a lateral da face, tanto do lado direito quanto do esquerdo. Estas imagens devem ser capturadas com as faces rotacionadas em aproximadamente 90° para cada lado. Outras imagens laterais com menor inclinação também podem ser capturadas, com aproximadamente 45° para cada lado. Outras poses são com a face inclinada para baixo e para cima, sendo que a inclinação deve estar entre 20 e 45°.

A expressão da face dos indivíduos de ser neutra, com os olhos abertos e a boca fechada, já que a expressão pode influenciar diretamente no desempenho de sistemas automatizados para reconhecimento de faces. Alguns fatores devem ser observados para que a expressão seja o mais neutra possível. Como exemplo de expressões não recomendadas tem-se: sorriso onde a parte de dentro da boca está exposta, sobrancelhas erguidas e sorriso intencional, entre outros (ISLAM, 2007; GRIFFIN, 2003).

Segundo as normas, não é aconselhável que o indivíduo esteja com acessórios que possam influenciar na imagem como, por exemplo: chapéu, brincos, colares, entre outros. Maquiagem, barba e bigode são características comuns que podem ser observadas nas pessoas. Por isso, é interessante alguns exemplos onde apareçam estes itens, para uma maior diversidade de características na base. Apesar das roupas aparecerem em todas as fotos, não existe uma restrição em relação as características que devem apresentar.

As imagens sendo capturadas devem, preferencialmente, ser centralizadas como mostra a Figura 15. Fazer isso no momento da captura facilita a definição de ROIs. A boca e o nariz devem estar centralizados, como mostrado na linha 'AA' da figura. O centro dos olhos deve estar numa linha horizontal imaginária (linha superior a marcação 'BB'), localizada aproximadamente a uma distância equivalente a 55% da altura da imagem, a partir da borda inferior. A largura da face deve ser de aproximadamente metade da largura da imagem, conforme ilustram as linhas 'CC' da Figura 15. (ISLAM, 2007; GRIFFIN, 2003; NIST, 1997)



**Figura 15: Imagem Frontal Centralizada.**

**Fonte: (ISLAM, 2007).**

#### 2.4.1.2 CONFIGURAÇÃO DO AMBIENTE

Uma das condições que deve ser tratada com cuidado durante a captura de imagens é a iluminação. Quando inadequada, poderá influenciar fortemente na extração de características

que, por sua vez, afeta o processo de reconhecimento. A iluminação deve ser igualmente distribuída em cada lado, em cima e embaixo da face, sendo possível observar todos os componentes de forma clara e livre de sombras. O aconselhado é que existam três fontes de iluminação para que esta seja bem distribuída pelo ambiente de captura (NIST, 2007), auxiliando na visualização da íris e da pupila e buscando evitar o *olho-vermelho*. Sombras ou muito brilho no plano de fundo devem ser evitados com uma iluminação adequada. Deve ser observado se a nada está sendo refletido em casos em que o indivíduo usa óculos (ISLAM, 2007). Nenhum documento especifica o tipo e a potência da lâmpada a ser utilizada na iluminação do ambiente.

O plano de fundo para captura de imagens deve ser liso e deve completar totalmente a imagem onde não está a face. A cor lisa recomendada pelo NIST (2007) é o cinza com uma intensidade de 18%, por ser uma cor com pouco brilho e ter pouca reflexão. Esta é representada pelo valor 118 para os 3 componentes do espaço RGB (*Red Green Blue*). Com um plano de fundo com estas características é possível estabelecer um melhor balanço das cores, refletindo as cores do espectro luminoso de forma uniforme. Deve estar sempre limpo, sem a presença de outros objetos, pois a diferença entre a pessoa e o fundo deve ser bem visível. Além destes cuidados, o indivíduo deve se posicionar a aproximadamente 70cm de distância, evitando o surgimento de qualquer sombra, assegurando que o plano de fundo fique uniforme (ISLAM, 2007).

#### 2.4.1.3 CONFIGURAÇÃO DA CÂMERA E DA IMAGEM

Islam (2007) também estabelece normas para configurar câmeras digitais, estas configurações estão resumidas na Tabela 1. As imagens faciais devem estar sempre em foco, facilitando o processo de identificação. A distância focal ajuda a minimizar a distorção da imagem e deve ser entre 70 e 100mm. A exposição deve ser regulada para que a saturação da imagem não seja alterada. A velocidade do obturador deve ser rápida o suficiente para que as imagens não fiquem borradas. A sensibilidade ISO da câmera deve ser o menor possível, preferencialmente um valor de 200 ou menos, pois isso reduz o ruído eletrônico produzido pelo sensor da câmera. A qualidade de cores deve ser configurada para que estas sejam capturadas com precisão. Em algumas câmeras, existe a configuração de cores, mas sua qualidade pode ser calibrada utilizando o plano de fundo adequado (ISLAM, 2007).

Na Tabela 1 ainda observa-se que as imagens devem ter no mínimo 480x600 pixels. Porém, quanto maiores, mais características podem ser extraídas. Deste modo, as imagens ficam com qualidade suficiente para serem utilizadas em sistemas de reconhecimento facial,

por exemplo. A relação Largura:Altura das imagens deve ser de 1:1,25 (NIST, 1997). Porém, para alguns casos, existem outras relações, tal como para imagens de passaporte (1:1,28) e para imagens de carteira de motorista (1:1,33). O espaço de cores recomendado é o RGB 24-bit.

**Tabela 1: Configurações da câmera e da imagem aconselhadas pelas normas internacionais.**

Distância Focal	70 a 100mm
Exposição	fixa
ISO	200 ou menos
Tamanho mínimo	480x600
Relação Largura:Altura	1:1,25
Espaço de Cores	RGB 24-bit
Flash	desativado

**Fonte: Autoria Própria**

#### 2.4.2 EXEMPLOS DE BASES DE IMAGENS FACIAIS

Atualmente está disponível na internet uma variedade de bases de imagens faciais para realizar o reconhecimento e outras tarefas relacionadas. Algumas delas estão resumidas na Tabela 2.

**Tabela 2: Comparativo das Características de Bases de Imagens Faciais.**

Base de Imagens	Pessoas	Total de Imagens	Resolução	Poses
FERET <sup>3</sup>	1199	30000	256x384	20
CVL Face <sup>4</sup>	114	798	640x480	7
PIE <sup>5</sup>	68	41368	640x486	13
BioID <sup>6</sup>	23	1521	384x286	—
M2VTS <sup>7</sup>	295	2460	720x576	—
Caltech <sup>8</sup>	27	450	896x592	—
Yale <sup>9</sup>	10	5760	640x480	9
MUCT <sup>10</sup>	276	3755	480x640	5
FEI <sup>11</sup>	200	2800	640x480	14

**Fonte: Autoria própria.**

<sup>3</sup><http://www.itl.nist.gov/iad/humanid/feret/>

<sup>4</sup><http://www.lrv.fri.uni-lj.si/facedb.html>

<sup>5</sup><http://www.ri.cmu.edu/>

<sup>6</sup><http://www.bioid.com/download-center/software/bioid-face-database.html>

<sup>7</sup><http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb>

<sup>8</sup><http://www.vision.caltech.edu/html-files/archive.html>

<sup>9</sup><http://cvc.yale.edu/projects/yalefaces/yalefaces.html>

<sup>10</sup><http://www.milbo.org/muct/>

<sup>11</sup><http://fei.edu.br/cet/facedatabase.html>

Dentre elas, a FERET é comumente utilizada em estudos sobre RF. Esta base foi utilizada em artigos científicos relevantes, como (PHILIPS; RIZVI, 2000), (RIZVI et al., 1998) e (WISKOTT et al., 1997). Na Figura 16 são mostrados alguns exemplos de imagens desta base.



**Figura 16: Exemplo de imagens da base FERET utilizadas em Philips e Rizvi (2000).**

**Fonte: (PHILIPS; RIZVI, 2000)**

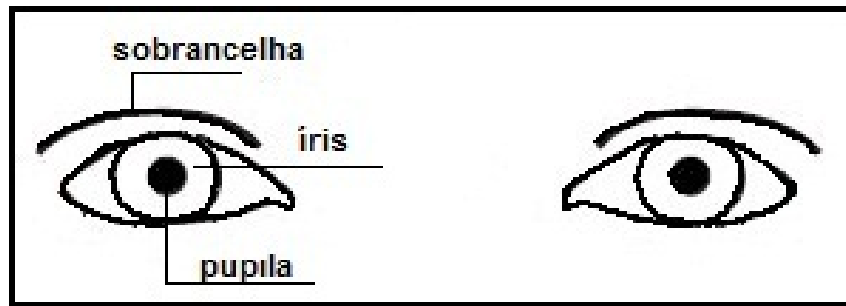
Nas demais, o número de pessoas varia entre 23 e 295 e o número total de imagens varia entre 450 e 41368. As únicas bases em que foram tiradas exatamente o mesmo número de fotos por pessoa são as bases da FEI e da CVL. Em outras, o número de fotos por pessoa varia e nem sempre é igual ao número de poses. As bases que informam o número de homens e mulheres que participaram da captura são: a CVL com 108 homens e 6 mulheres e a FEI com 100 homens e 100 mulheres. Um diferencial entre todas as bases da tabela é a da MUCT, que possui 76 pontos de marcação na face, chamadas de pontos fiduciais, sendo que nenhuma outra base apresenta estes pontos. Estas marcações são utilizadas no processo de RF. A única base pública brasileira existente até a execução deste trabalho é a base da FEI (OLIVEIRA; THOMAZ, 2005), que conta com 14 imagens por pessoa totalizando 2800 imagens. Todas as bases observadas possuem baixa resolução, com as imagens menores que do 1 megapixel.

## 2.5 DETECÇÃO DE REGIÕES DE INTERESSE

A detecção de RoIs é importante para o RF, pois é a partir delas que se iniciam outros processos, como a normalização e a extração de características. Algumas possibilidades são a face, a região dos olhos e o centro do olho (pupila) (Figura 17). No caso deste trabalho, a RoI mais importante é a pupila, utilizada na normalização.

A detecção de olhos é responsável por encontrar a posição dos olhos em uma face. Muitas vezes, esta posição corresponde a região onde estão os dois olhos, ou ainda, a uma localização mais precisa, que é o ponto central de cada um dos olhos. A região dos olhos pode ser representada através dos dois olhos e, algumas vezes, as sobrancelhas, como mostrada na Figura 17. Já o ponto central é representado pela localização da pupila (Figura 17).





**Figura 17: Região dos Olhos.**

**Fonte: Autoria Própria.**

Os dois principais métodos de detecção de olhos são o ativo e o passivo (LEITE, 2008).

O método ativo é usado em sequências de vídeos. Uma das formas de detectar os olhos neste método é através da piscada de olhos, que é analisada através da diferença de vários *frames* do vídeo. Outra forma é através de iluminação infravermelha, sendo um feixe direcionado sobre os olhos. Com isso, a córnea reflete a luz produzindo o efeito dos "olhos vermelhos" e este efeito faz com que as pupilas sejam regiões de alto brilho. Assim, é possível determinar a posição do olho através destas regiões. Uma desvantagem em utilizar métodos ativos é que as imagens devem ser capturadas com alguns requisitos como, por exemplo, aparelhagem especial para emissão de luz infravermelha, sem que ela atrapalhe a captura.

O método passivo é mais utilizado para a detecção de olhos em imagens estáticas e possui três abordagens: análise de aparência, aprendizagem e modelos (LEITE, 2008).

Na primeira abordagem, análise de aparência, são utilizadas características de cor, textura e forma. Agrupamentos de pele são formados por um espaço de cor, feitos através de análise da pele humana presente nas imagens. A imagem é binarizada utilizando os limites destes agrupamentos. Em seguida, são aplicadas regras de conhecimento para determinar a real posição dos olhos. As vantagens desta abordagem são a necessidade de poucos recursos computacionais e, também, o fato de os cálculos das características serem invariantes a escala, rotação, cores ou formas. A principal desvantagem é a dificuldade de estabelecer critérios para identificar pele de várias etnias e em imagens com variação de iluminação (LEITE, 2008).

Para a abordagem de aprendizagem, um classificador é treinado utilizando um conjunto de imagens de exemplos e de contraexemplos. O treinamento pode ser feito utilizando técnicas como Redes Neurais e SVMs, entre outras. Nesta abordagem, os problemas de rotação, escala e iluminação podem ser anulados pela geração de vários exemplos que representem estas características. A maior desvantagem está em definir o que é um bom contraexemplo para o

treinamento (LEITE, 2008).

A detecção por modelos tem como ideia o casamento de uma imagem modelo com regiões da imagem com face, como exemplo o *Template Matching*. Uma possível imagem modelo é a de imagens de olhos recortadas de uma face, outra forma é através de um padrão composto por um círculo e duas curvas, representando a morfologia de um olho (LEITE, 2008). A principal desvantagem deste método é o alto custo computacional, pelo grande espaço de busca e pela necessidade de modificar o modelo em relação a escala e rotação.

Para este trabalho foi utilizada a abordagem por aprendizagem, em duas metodologias distintas. Uma utiliza SVM e outra AdaBoost, algoritmos de reconhecimento de padrões. A detecção de ROI é utilizada como etapa inicial da normalização automática neste trabalho, mas também pode ser utilizada para localizar pontos de interesse para extração de características.

## 2.6 NORMALIZAÇÃO DE IMAGENS FACIAIS

A normalização de imagens é uma etapa que visa auxiliar diversas áreas da visão computacional. Esta etapa pode ser considerada como um pré-processamento de imagens e que tem como objetivo padronizá-las. As principais características ajustadas no processo de normalização são a rotação e o redimensionamento das imagens, posicionamento de ROIs e centralização da imagem. Estas transformações são classificadas como *Affine Transformation*, quando aplicadas em onde um determinado conjunto de pixels situado em uma região da imagem original se mantém na mesma posição na imagem final. Outra característica mantida é a relação entre as medidas de distância. Por exemplo, a proporção entre a distância dos olhos e a distância do ponto central dos olhos com o nariz são mantidas, apesar das distâncias em si diminuírem ou aumentarem. Matematicamente, estas transformações utilizadas na normalização facial, podem ser consideradas transformações lineares.

Além destas transformações, existem outras modificações que podem ser realizadas nas imagens, como por exemplo, padronização de cor de fundo, brilho e iluminação. Todas estas características são alteradas com intuito de manter um padrão entre todas as imagens de uma base.

O RF utiliza a etapa de normalização para padronizar as imagens visando facilitar as etapas seguintes, tal como a extração de características. O principal passo da normalização de imagens faciais é encontrar os pontos centrais dos olhos, ou seja, as coordenadas de cada pupila. Estes dois pontos são essenciais para realizar grande parte das etapas de normalização.

Assim como na captura de imagens faciais, para realizar a normalização de imagens

faciais existem algumas normas internacionais que devem ser seguidas.

### 2.6.1 NORMAS PARA NORMALIZAÇÃO DE IMAGENS FACIAIS

Um dos fatores que exige o processo de normalização das faces é que, muitas vezes, a face frontal pode estar com rotação maior do que a permitida pelas normas de captura, que é de  $\pm 5^\circ$ . Outro fator é a relação Largura:Altura, que apresenta orientação vertical nas normas de captura e orientação horizontal em grande parte das câmeras (ISLAM, 2007).

O principal documento que discute algumas regras para normalização de imagens faciais é Griffin (2003). As normas estabelecidas são baseadas nas normas de captura de imagens da NIST (1997) e são descritas a seguir.

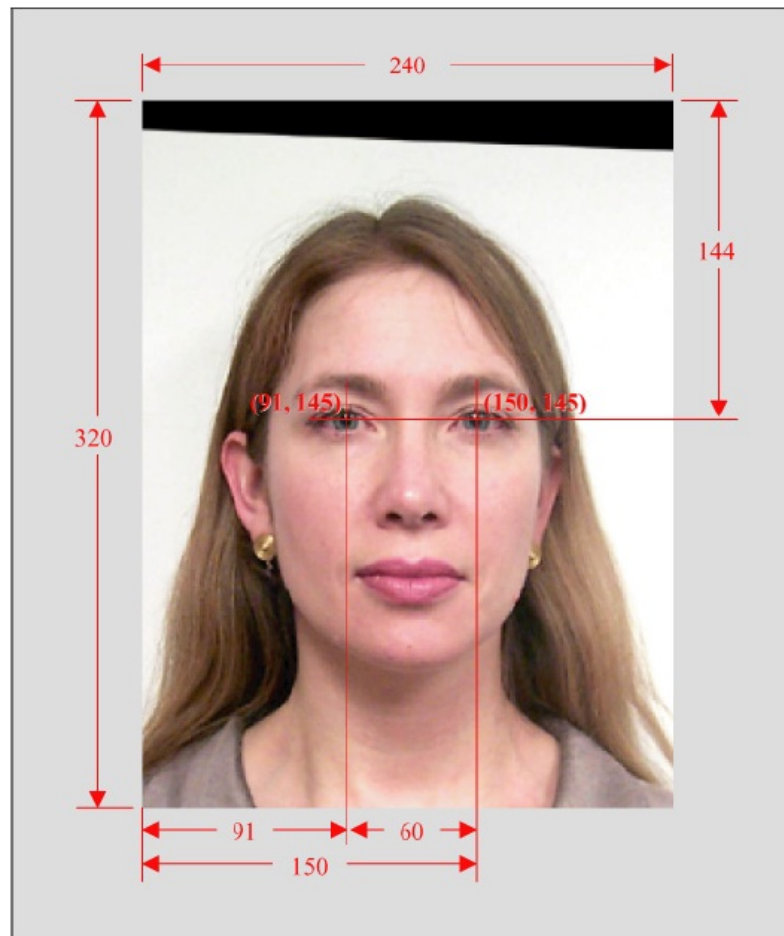
A Figura 18 mostra a imagem normalizada canônica, de tamanho de 240x320 pixels. O tamanho da imagem normalizada é menor do que da imagem capturada (que, segundo as normas, deve ser de 480x600). Isso ocorre pois elementos menos importantes, como parte do fundo ou dos ombros da pessoa, são recortados e a imagem original não deve ser ampliada, pois isso pode alterar sua qualidade.

O primeiro passo da normalização é a determinação da posição dos olhos. Os olhos de todas as faces devem estar localizados sempre nos mesmos pontos da imagem, (91, 145) e (150, 145), a uma distância de 60 pixels um do outro. Na imagem, é possível notar uma região preta no alto da imagem. São pixels adicionados para que a imagem ficasse do tamanho correto, já que a imagem não pode ser ampliada. (GRIFFIN, 2003)

Estas especificações foram escolhidas por algumas razões. A medida de 60 pixels entre cada olho é uma distância suficiente para descrever características faciais para reconhecimento e verificação facial. Além disto, o posicionamento fixo dos olhos facilita a extração de ROI. A parte de fundo restante na imagem é utilizada como margem de erro, caso o olho seja colocado em um lugar errado ou quando uma face é maior.

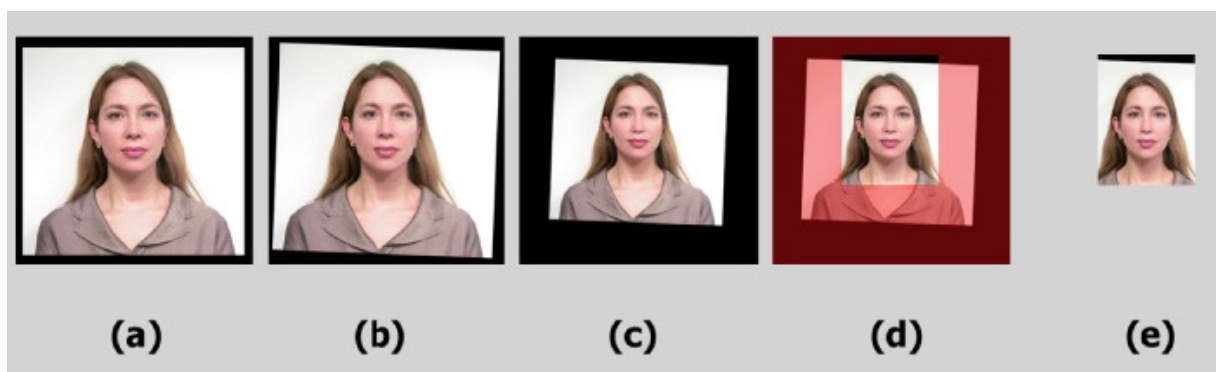
A imagem normalizada é criada através de *Affine Transformation*, que mapeia as coordenadas dos olhos para as coordenadas da imagem canônica. O processo de *Affine Transformation* consiste em rotação, mudança de escala e translação da imagem. A imagem final deve então ser recortada para o tamanho de 240x320 pixels. A Figura 19 mostra o processo de normalização de uma imagem facial.

Na Figura 19 pode-se notar a imagem original (a) e as etapas da normalização após a definição dos pontos centrais dos olhos. Na Figura 19 (b) é mostrada a face rotacionada. Então a imagem é redimensionada para obter 60 pixels entre os olhos Figura 19 (c). Por fim, os olhos



**Figura 18: Medidas da Imagem Frontal Normalizada Canônica (em pixels).**

**Fonte: (GRIFFIN, 2003).**



**Figura 19: Processo de Normalização de uma Imagem Facial.**

**Fonte: (GRIFFIN, 2003).**

são posicionados nas coordenadas corretas Figura 19 (d) e então a imagem é recortada Figura 19 (e). (GRIFFIN, 2003)

## 2.7 TRABALHOS CORRELATOS

No trabalho de Jian e Honglian (2009), a detecção de olhos foi feita seguindo uma série de etapas. Primeiramente foi selecionada a região da face através da cor da pele. Utilizando apenas o componente vermelho, foi feita a segmentação para extrair a face. Em seguida, somente na região da face, foi realizada a detecção de bordas e aplicadas operações morfológicas, principalmente o fechamento e a erosão. Na etapa seguinte, foram calculados área, distância e ângulo entre os olhos, resultando em algumas regiões candidatas a olhos. Finalmente, foram estabelecidos diversos *templates* de olhos com variações de ângulos. Através do *Template Matching* (GONZALEZ; WOODS, 2006), os *templates* foram aplicados nas regiões candidatas a olhos para efetivamente realizar a sua detecção. Para testes foram utilizadas 200 imagens, sendo que o tamanho da região detectada da face não ultrapassava 80x80pixels. A melhor taxa de acerto foi de 91,23% para faces com rotação menor que 20° e o pior caso foi para rotações maiores que 60° onde a taxa de acerto foi de 62,21%.

Em Ke e Huang (2009), também foi utilizada a detecção da região da face através da cor da pele, mas utilizando o espaço de cores YCrCb. Através do filtro de Gabor foram detectadas ROIs dos olhos. Utilizando operações morfológicas foi reduzido o número de regiões candidatas a olhos. Através de dois *templates* obtidos através da média de 20 imagens de olhos, direito e esquerdo, foi realizado o *Template Matching* (GONZALEZ; WOODS, 2006) com as regiões candidatas. Por último, usando a distância Euclidiana, foi definido se os candidatos eram olhos esquerdo ou direito. Ke e Huang (2009) obtiveram 94% de acerto com imagens frontais e 83% para imagens com rotação, utilizando a base de imagens FERET (NIST, 2004).

imagens em níveis de cinza com baixa resolução também foram utilizadas no estudo de Jacques et al. (2009). Neste trabalho, após a detecção da face, foram detectadas duas regiões, uma para cada olho, utilizando o algoritmo Viola e Jones (VIOLA; JONES, 2001). A posição da pupila de cada olho foi então localizada através do *Template Matching* (GONZALEZ; WOODS, 2006), utilizando soma das diferenças absolutas. Jacques et al. (2009) perceberam que, quando a imagem da face é transformada para níveis de cinza, a região central do olho é formada por um círculo preto, e assim quatro *templates* sintetizados foram definidos. Os testes foram realizados em quatro bases de imagens: BioID (FRISCHHOLZ; DIECKMANN, 2000 apud JACQUES et al., 2009), Caltech (WEBER, 1999 apud JACQUES et al., 2009), IMM (NORDSTROM et al., 2004 apud JACQUES et al., 2009) e *Japanese Female* (AKAMATSU et al., 1998 apud JACQUES et al., 2009) Através deles, a melhor taxa média de acerto, em torno de 95,5%, foi obtida com o *template* com um círculo preto e uma pequena borda branca.

Para determinar regiões candidatas, Jesorsky et al. (2001) utilizaram a segmentação da imagem através da detecção de bordas seguida da aplicação de um *threshold* local para binarização. A imagem resultante é comparada a um modelo com a região candidata utilizando a Distância de Hausdorff Modificada (DUBUISSON; JAIN, 1994 apud JESORSKY et al., 2001). Para melhorar a precisão da detecção, Jesorsky et al. (2001) utilizaram uma rede neural MLP treinada com as pupilas centralizadas na imagem. O melhor resultado obtido foi de 98,4% utilizando a base M2VTS (CHAM, 2006).

No trabalho de Prodossimo et al. (2011) foi utilizada a técnica de *Template Matching* (GONZALEZ; WOODS, 2006). As imagens utilizadas nos experimentos não passaram por nenhum pré-processamento. Os *templates* utilizados nos experimentos foram obtidos das próprias imagens faciais. A busca para o *matching* dos *templates* foi otimizada com dois algoritmos de otimização, Colônia de Abelhas (*Artificial Bee Colony - ABC*), utilizado por Chidambaram e Lopes (2010), e Busca Harmônica (*Harmony Search - HS*), utilizado por Scalabrin et al. (2010). O resultado deste estudo comparativo mostrou que o ABC foi mais eficiente que o HS, devido ao fato do *template* ser gerado da própria face em que se deseja encontrar o olho. Este estudo comparativo foi baseado nos trabalhos do Perlin (PERLIN et al., 2008; PERLIN, 2009) que utilizaram PSO (*Particle Swarm Optimization*) para o reconhecimento de objetos e (CHIDAMBARAM; LOPES, 2009), que utilizaram o ABC para otimizar o processo de reconhecimento de objetos presentes num imagem de paisagem usando o *Template Matching*. Da mesma forma, o ABC também foi utilizado em estudos de detecção de íris por (JACQUES et al., 2009).

A detecção de sorrisos no trabalho de (FREIRE et al., 2009) utiliza uma combinação do LBP, para extração de características, com o SVM, para classificação das imagens. Para o treinamento da SVM foram utilizadas imagens de 59x65 pixels, sendo que existiam 2421 imagens de faces com sorriso e 3360 sem sorriso. Nos melhores casos a taxa de acerto do algoritmo de (FREIRE et al., 2009) foi de aproximadamente 90%.

Uma proposta de captura de imagens foi realizada por Prodossimo et al. (2012). Neste trabalho, foram capturadas um total de 4050 imagens no tamanho de 2592x1944 pixels. Dentre estas, as frontais passaram pelo processo de normalização, realizada com a escolha dos pontos centrais dos olhos de forma manual. O restante do processo foi realizado de forma automática. As imagens normalizadas passaram a ter um tamanho de 1200x1500 pixels.

No trabalho de Oliveira e Thomaz (2005), o processo de normalização foi citado como alinhamento de imagens, mas somente o alinhamento de olhos foi realizado. Ao mesmo tempo, o processo foi realizado de forma manual, baseado num programa comercial. Após o alinhamento de imagens, estas foram recortadas para o tamanho de 260 x 360 pixels.

Ao contrário de Oliveira e Thomaz (2005), o processo de normalização de imagens foi feito de forma automática em Amaral e Thomaz (2008). Porém, a definição dos pontos centrais do olho era de forma manual. Neste trabalho foram realizadas a rotação para alinhar os olhos, escala, recorte, descoloração e suavização de contorno. A escala foi ajustada para que a distância entre os olhos ficasse com 100 pixels. As imagens foram recortadas e ficaram com tamanho de 250 x 300 pixels. A etapa de descoloração consiste em alterar as imagens para níveis de cinza e tem como objetivo diminuir o esforço computacional (AMARAL; THOMAZ, 2008). Durante a normalização, a rotação gerou um efeito serrilhado na imagem que, por sua vez, forçou a suavização dos contornos. A base de imagens utilizada para os experimentos foi a FEI *Face Database* (OLIVEIRA; THOMAZ, 2005).

O trabalho de Amaral e Thomaz (2008) foi ampliado em (AMARAL et al., 2009) para normalização em ambientes controlados e não-controlados. As etapas de normalização como a rotação, a escala, o recorte, a descoloração e a suavização de contorno foram realizadas do mesmo modo como no trabalho anterior. Porém, uma etapa adicional chamada Equalização de Histograma foi feita neste trabalho. Esta etapa é responsável por melhorar o contraste das imagens em níveis de cinza. Duas bases foram utilizadas: uma controlada, FEI *Face Database* (OLIVEIRA; THOMAZ, 2005), e outra não-controlada, base pública da Redesap (Rede Nacional de Identificação e Localização de Crianças e Adolescentes Desaparecidos).

As técnicas apresentadas neste capítulo são base para o desenvolvimento da metodologia proposta no próximo capítulo (Capítulo 3). Construção de uma base de imagens, normalização das imagens e detecção de ROI são as principais etapas discutidas, sendo que as duas primeiras foram desenvolvidas baseadas nas normas internacionais mencionadas anteriormente.

### 3 DESENVOLVIMENTO

O desenvolvimento deste trabalho foi dividido em três áreas principais, correspondentes a etapas do RF: construção de uma base de imagens faciais, detecção de ROI e normalização facial. No decorrer deste capítulo estão dispostas as metodologias para execução destas três etapas, seguindo as normas e utilizando as técnicas citadas no capítulo anterior.

#### 3.1 CAPTURA DAS IMAGENS FACIAIS

Levando em consideração as normas estabelecidas pelos documentos ((ISLAM, 2007), (NIST, 2007), (FACIAL IDENTIFICATION SCIENTIFIC WORKING GROUP, 2011) e (GRIF-FIN, 2003)) citados no Capítulo 2 foi construída uma base de imagem faciais. Além de levar em conta as normas destes documentos, procedimentos foram modificados de acordo com a necessidade deste e de outros trabalhos correlacionados ao RF. Porém, os novos procedimentos ainda são relacionados com as normas especificadas nos documentos citados.

##### 3.1.1 EQUIPAMENTOS

A captura de imagens foi realizada utilizando-se uma câmera digital posicionada sobre um tripé para que a distância entre a câmera e as pessoas fosse mantida. A resolução de cada foto é de 2592x1944 pixels (5 megapixels). Esta resolução foi escolhida por ser maior do que as existentes na literatura e porque permite ajustes posteriores quanto ao tamanho e à resolução, ou seja, será possível processar a imagem capturada para obter um tamanho menor, sem perda de informações. Quando a resolução é pequena, não é possível aumentar a imagem, pois ocorre uma perda na qualidade da imagem, interferindo nas informações e características. Com essa resolução é possível observar que a relação Largura:Altura é de 1,33:1 (Tabela 3) da imagem está diferente do que é solicitado pelas normas (Tabela 1). Porém, isto será ajustado a medida que as imagens forem normalizadas.

A câmera utilizada foi uma Sony Cyber-Shot (DSC-S3000) de 10.1 Megapixels, mas



as imagens foram capturadas com somente 5 Megapixels, suficiente para a resolução desejada. Para todas as posições de face, a câmera estava com um valor de ampliação de três vezes (*zoom* de 3x). A ampliação foi necessária para obter somente a face da pessoa na imagem capturada, mantendo a distância adequada em relação à face, a fim de, garantir melhor nitidez e evitar o aparecimento de sombras. Outras características podem ser observadas na Tabela 3, seguindo as configurações estabelecidas pelas normas mostradas anteriormente na Tabela 1. Porém limitando-se as configurações da câmera utilizada, como é o caso da distância focal que é de 12mm enquanto as normas solicitavam uma distância de 70mm a 100mm.

**Tabela 3: Configurações da câmera e da imagem utilizadas.**

Distância Focal	12mm
Exposição	1/80s
ISO	200
Tamanho mínimo	2592x1944
Relação Largura:Altura	1,33:1
Espaço de Cores	sRGB 24-bit
Flash	desativado

**Fonte: Autoria Própria**

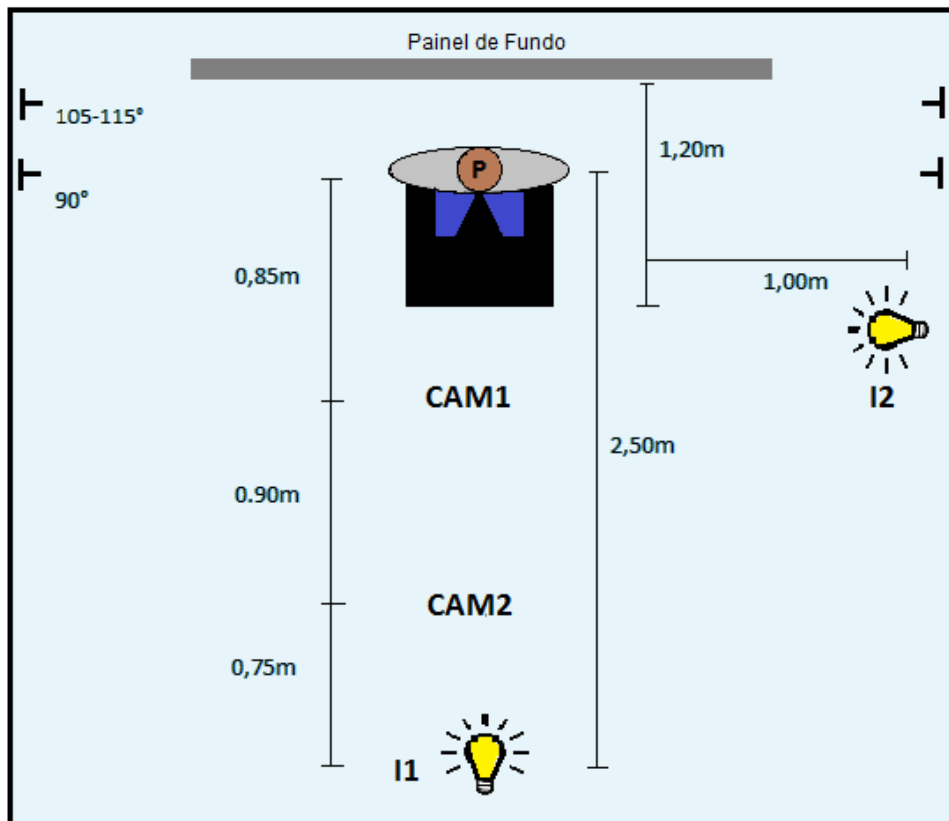
Um painel de cor fixa foi confeccionado para ser utilizado como plano de fundo no ambiente de captura. A cor do painel é o cinza 18% (NIST, 2007), conforme foi aconselhado nas normas descritas no capítulo anterior. Com isto, pôde-se melhorar o controle de qualidade das imagens e, também, distinguir melhor a diferença entre a pessoa e o fundo.

Pontos fixos foram definidos no chão para que o tripé, cadeira e iluminação fossem posicionados sempre no mesmo lugar. Com isto, foi possível manter um controle da distância entre a câmera e a pessoa, assim como da distância entre a iluminação e os outros componentes. Deste modo, as imagens puderam ser capturadas mantendo-se padrão, buscando manter a qualidade.

A iluminação do ambiente de captura das imagens foi feita artificialmente e de duas maneiras. Algumas imagens foram capturadas com fontes luminosas próprias para fotografar faces e outras com luzes ambiente da sala, caracterizando duas formas distintas de iluminação, ambas consideradas controladas. Neste trabalho o termo controlada é utilizado para indicar que não existe variação na intensidade e na cor da luz e não há interferência de iluminação externa, ou seja, independente do horário elas serão sempre a mesma. Além disto, foram capturadas imagens com a incidência lateral de iluminação, com objetivo de gerar sombra em um dos lados da face.

### 3.1.2 POSICIONAMENTO DO EQUIPAMENTO

Com base nos estudos de Gross (2005), de Rubinfeld e Wilson (1999) e de Islam (2007), relacionados às normas de captura de imagens, os equipamentos utilizados neste estudo foram posicionados como mostrado na Figura 20. São apresentadas as posições do tripé, assim como o posicionamento da cadeira, plano de fundo, marcações e fontes de iluminação. Na Figura 20 os símbolos marcados com 'I1' e 'I2' são as iluminações próprias para a captura de imagens. A letra 'P' representa o local onde fica a pessoa. As marcações 'CAM1' e 'CAM2' são as duas posições do tripé/câmera e as marcas laterais (90° e 105-115°) são marcações para o voluntário olhar durante a captura das imagens laterais.



**Figura 20: Distâncias e Posições do Equipamento para Captura de Imagens.**

**Fonte: Autoria própria**

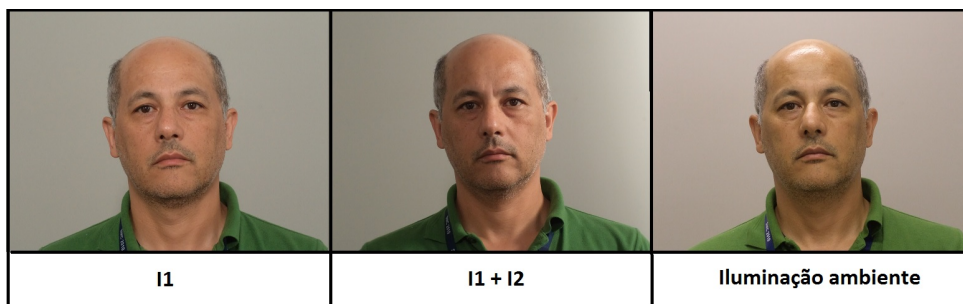
Primeiramente, a cadeira foi posicionada um pouco à frente do painel de fundo para evitar sombras. Os pés da frente da cadeira estão a 1,2 metros do painel, fazendo com que a pessoa fique a aproximadamente 70 centímetros do fundo (ponto 'P' da Figura 20), como aconselhado no trabalho de Islam (2007).

A Figura 20 também apresenta as marcações da câmera. Primeiramente, posicionada

a 0,85 metros do centro da cadeira ('CAM1'), de modo a aparecer somente a face e uma parte do ombro da pessoa na imagem adquirida. Posteriormente, a câmera foi colocada a 1,75 metros do fundo da cadeira ('CAM2'), para algumas fotos em escala. A variação das distâncias foi escolhida para ter uma maior variabilidade de imagens para serem utilizadas nos estudos de RF.

Com o objetivo de capturar as imagens com variações de iluminação, foram utilizadas duas fontes de iluminação próprias, além da iluminação original do ambiente da sala. As duas fontes de iluminação própria, 'I1' e 'I2' na Figura 20, são equipadas com lâmpadas halógenas com uma potência de 1000 Watts cada e as luzes ambientes da sala são fluorescentes com potência de 32 Watts. As fontes 'I1' e 'I2' garantem uma iluminação uniforme por serem próprias para captura de imagens de faces, já a iluminação ambiente não é uniforme, porém é interessante para a variabilidade de imagens na base.

A iluminação 'I1' foi posicionada a 2,5 metros do fundo da cadeira, de forma centralizada com a face a uma altura de aproximadamente 2 metros. Esta posição, junto com a distância da cadeira com o plano de fundo, evitaram que aparecessem sombras no painel. A outra fonte, 'I2', foi localizada na lateral da cadeira, a 1 metro de distância e 1,7 metros de altura. Quando está ligada, a segunda fonte luminosa ocasiona sombra em um dos lados da face. Neste estudo, este procedimento foi utilizado para causar uma variação significativa de iluminação na imagem, fazendo com que um lado da face esteja bem iluminado e o outro com sombra. Os efeitos da iluminação para as três variações pode ser visto na Figura 21.



**Figura 21: Efeitos das diferentes formas de Iluminação.**

**Fonte: Autoria própria**

Além dos pontos fixos que foram definidos no chão, como citado anteriormente, mais duas marcas foram definidas para as fotos laterais (90° e 105-115°). Estas marcações foram feitas nas paredes laterais da sala de captura, para onde a pessoa deve focalizar mantendo os olhos centralizados. De fato, estes pontos serviram para auxiliar na captura da lateral da face.

Para as fotos com a face voltada para cima foi definido um ponto junto à luz artificial para que as pessoas olhassem sempre para o mesmo ponto. Para as fotos com a face voltada

para baixo outro ponto foi definido em um dos pés do tripé. Estes pontos asseguram que o ângulo de inclinação não varie muito, existindo uma pequena mudança dependendo da altura da pessoa.

Apesar dos vários pontos de controle, durante o processo de captura o voluntário deve ser instruído para que melhore sua pose, de acordo com a visão de quem está realizando a captura da imagem. Além das inclinações e rotações, o voluntário era informado sobre o posicionamento dos seus olhos, de modo que estivesse de acordo com cada pose. As poses das faces, assim como o procedimento adotado para a captura estão descritos a seguir.

### 3.1.3 POSES DAS FACES

As imagens foram capturadas com variações de iluminação, todas controladas, utilizando iluminação artificial própria para captura de imagens faciais e a iluminação original da sala, sendo estas variações divididas em cenários, como mencionado a seguir:

1. cenário C1 - fonte de iluminação I1 (própria para fotos de face) ligada e a iluminação ambiente desligada;
2. cenário C2 - cenário C1 mais a fonte I2 ligada, com o objetivo de gerar sombra lateral na face;
3. cenário C3 - iluminação controlada com as luzes ambiente ligadas e as fontes próprias ('I1' e 'I2') desligadas.

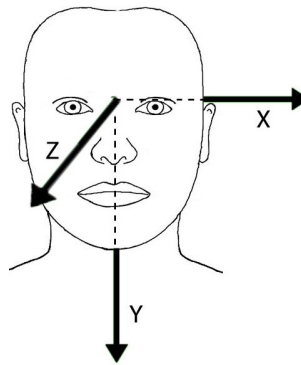
Todas as fotos foram tiradas com o mesmo painel de fundo, 18% cinza. A maioria das fotos foi tirada sem expressão facial, para cada voluntário foi solicitado a expressão de um sorriso para a captura de duas fotos, para uma maior variabilidade de imagens. Foi observado se não existia cabelo na testa ou, pelo menos, que o cabelo não ocultasse elementos relevantes do rosto, tais como sobrancelhas, olhos e orelhas. Para as fotos laterais a orelha é um ponto importante para extração de características em alguns algoritmos e, por isto, foi garantido que a mesma fosse bem visível. Neste trabalho foi utilizada para a normalização lateral (Seção 3.3.2).

O número total de imagens capturadas por pessoa neste estudo foi de 27, contando com todas as variações de iluminação e de pose. De modo geral, as imagens capturadas podem ser classificadas em duas categorias: 14 imagens com iluminação própria, sendo 13 no cenário C1 e 1 no C2, e 13 imagens com o cenário C3. Foram capturadas imagens frontais, frontais com inclinações, imagens laterais e outras imagens com a câmera em posições diferentes.

### 3.1.4 PROCEDIMENTO PARA CAPTURA DAS IMAGENS

Antes de iniciar a captura das imagens, num primeiro momento, os participantes deveriam assinar um Termo de Consentimento Livre e Esclarecido (Anexo A), concordando com a participação na pesquisa e autorizando o uso de suas imagens para fins científicos.

Com o objetivo de facilitar o processo de captura, foi estipulada uma sequência que pode ser observada na Tabela 4. As imagens foram capturadas com rotações nos eixos  $x$ ,  $y$  e  $z$  em relação às imagens frontais, estes eixos estão representados na Figura 22. Algumas imagens foram capturadas com translação no eixo  $z$ , ou seja, imagens em escala. Outras com inclinações para cima e para baixo, variando o eixo  $x$ . E as imagens laterais são consideradas rotações no eixo  $y$  da face. Grande parte das poses foram capturadas de acordo com as normas, porém algumas foram adaptadas de acordo com a necessidade do trabalho, estas estão marcadas com '\*' na Tabela 4



**Figura 22: Eixos de rotação e translação da face.**

**Fonte: Autoria Própria.**

A captura foi realizada em duas etapas, inicialmente o cenário C1 e depois C3 seguindo a sequência da Tabela 4. Aproveitando o cenário C1, já foi feita a captura da imagem 03 com a iluminação lateral (cenário C2). Esta imagem é codificada como número 14, por ser somente uma imagem para o cenário C2, variação do C1.

Nas imagens 01 e 02, capturou-se a face lateral esquerda do voluntário em 90 graus e com uma leve inclinação para direita em 105 a 115 graus, este caso foi adaptado das normas para esconder um pouco mais a face do voluntário. O mesmo foi aplicado para as imagens 09 e 10, desta vez com a face lateral direita. As imagens 03 até 06 são frontais, com variações de inclinação e expressão facial, poses exigidas pelas normas internacionais (Seção 2.4.1). Também de acordo com solicitado, a imagem 07 é capturada com a pessoa olhando para cima e a 08, olhando para baixo. As imagens 11, 12 e 13 são semelhantes às imagens 03, 05 e

**Tabela 4: Sequência de Captura das Imagens para cada cenário.**

Código	Posição
01	90° Esquerda (rotação no eixo Y)
02	* 105° Esquerda (rotação no eixo Y)
03	Frontal
04	Frontal com Expressão
05	Inclinação Direita (rotação no eixo Z)
06	Inclinação Esquerda (rotação no eixo Z)
07	Face voltada para Cima (rotação no eixo X)
08	Face voltada para Baixo (rotação no eixo X)
09	90° Direita (rotação no eixo Y)
10	* 105° Direita (rotação no eixo Y)
11	* 2,5m Frontal (translação em Z)
12	* 2,5m Inclinação Direita (translação e rotação no eixo Z)
13	* 2,5m Inclinação Esquerda (translação e rotação no eixo Z)

**Fonte: Autoria Própria.**

06, respectivamente, porém com a câmera a uma distância de 1,75 metros, a escala foi adotada para alterar a incidência de luz na face do indivíduo.

Contando com a utilização de iluminação própria para fotos, o flash da câmera foi desativado e o tempo do obturador da máquina, tempo de exposição, zoom entre outras características da câmera devem ser fixos, como apresentado na Tabela 3. Com isso é possível ter um maior controle de como as imagens estão sendo capturadas. Este controle é necessário para permitir comparações entre outras bases de dados ou entre algoritmos.

Outra forma de estabelecer um controle para comparações é a utilização de rotações e iluminação fixas. Por isso foram feitas as marcações no ambiente de captura, como relatado anteriormente.

#### 3.1.4.1 CODIFICAÇÃO DOS ARQUIVOS DE IMAGEM

De modo a facilitar a identificação das imagens, foi estabelecida uma codificação para as características de cada imagem. Esta codificação também garante o sigilo dos voluntários do estudo, pois suas identidades não são informadas. A codificação segue o formato "NNN-I-PP-S-F-A.jpg", onde cada código representa uma informação como apresentado na Tabela 5:

**Tabela 5: Codificação dos Arquivos de Imagem.**

Código	Informação
NNN	código da pessoa, variando de 001 até 138
I	tipo de iluminação/cenário utilizado, onde 1 é o cenário C1 e 2 é o C3 *o cenário C2 também utiliza código 1 para a pose 14, por ser uma variação do C1
PP	código da pose da face para captura, valores mostrados na Tabela 4
S	sexo, sendo 1 para Homens e 2 para Mulheres
F	faixa-etária do voluntário, códigos apresentados na Tabela 6
A	código auxiliar para futuras anotações

**Fonte: Autoria Própria**

**Tabela 6: Codificação da Faixa Etária.**

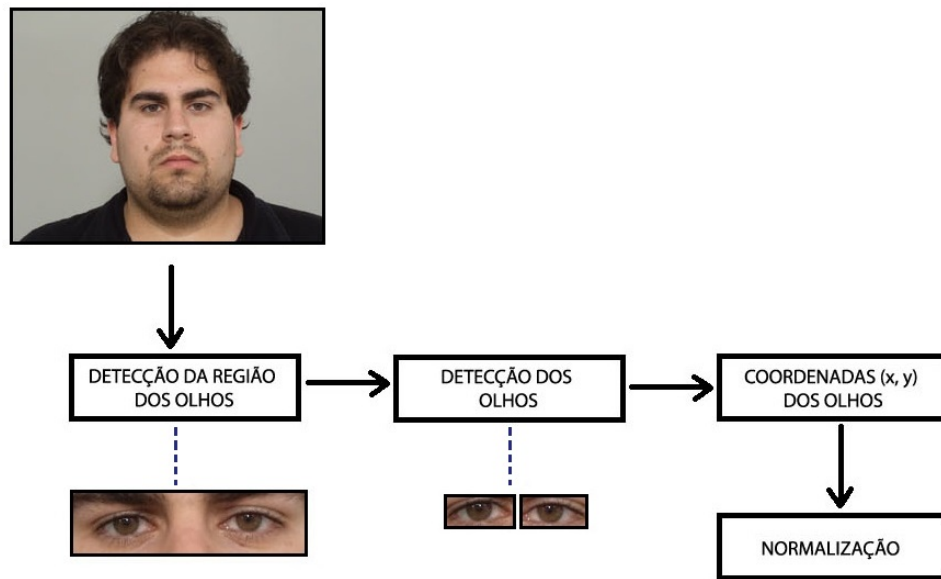
Código	Faixa Etária
0	idade não informada
1	menos que 20 anos
2	entre 21 e 30 anos
3	entre 31 e 40 anos
4	entre 41 e 50 anos
5	entre 51 e 60 anos
6	entre 61 e 70 anos

**Fonte: Autoria Própria.**

### 3.2 DETECÇÃO DE OLHOS

Foram aplicadas duas metodologias de detecção de olhos com o objetivo de comparação, tanto de porcentagem de acerto quanto de tempo de processamento. Primeiramente os olhos foram detectados com uma metodologia que utiliza *Haar-Like Features* e Adaboost. A estrutura é semelhante ao Viola e Jones, porém este algoritmo foi adaptado para identificar características referentes a regiões dos olhos e não a face inteira. A outra metodologia foi criada com base na anterior com a finalidade de testar a qualidade dos dois algoritmos. Neste caso as técnicas utilizadas foram o LBP e o SVM. Ou seja, o princípio das técnicas utilizadas em ambos algoritmos é a mesma, os dois utilizam um algoritmo para filtrar as imagens e, em seguida, um algoritmo de reconhecimento de padrões.

O procedimento de detecção dos olhos (Figura 23) utilizado foi o mesmo para as duas metodologias, diferindo somente pelas técnicas utilizadas. Primeiramente, a imagem é apresentada ao sistema, que detecta uma região em que podem estar situados os dois olhos. Em seguida, dentro desta região, o algoritmo busca os dois olhos. Finalmente, a coordenada central de cada olho é utilizada no processo de normalização, procedimento que é explicado na Seção 3.3.



**Figura 23: Etapas para detecção dos olhos.**

**Fonte: Autoria Própria.**

### 3.2.1 HAAR-LIKE FEATURES COM ADABOOST

Esta metodologia é baseada em *Haar-like Features* e Adaboost (HA). Diferente do algoritmo de Viola e Jones, as características Haar-like foram treinadas para detectar, primeiramente, a região compreendida pelos dois olhos e, em seguida, dentro desta região, encontrar um olho de cada vez de forma mais precisa.

O algoritmo age como uma janela deslizante, realizando recortes na imagem de face. Em cada um destes recortes são extraídas as *Haar-like Features* e então o recorte é classificado pelo algoritmo de reconhecimento de padrões Adaboost. Em um primeiro momento o classificador indica se a região a ser classificada contém os dois olhos do indivíduo. Após isto, dentro desta região é feita uma nova busca, agora tentando encontrar um olho de cada vez, para obter os pontos centrais necessários para a normalização.

### 3.2.2 LBP COM SVM

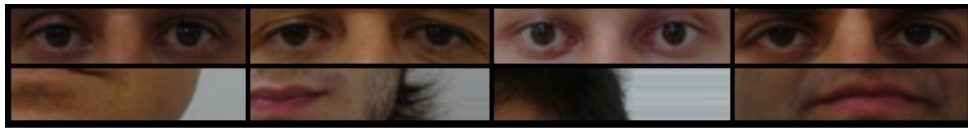
A ideia central desta metodologia é a mesma que a anterior, diferindo apenas pelas técnicas utilizadas. Nesta foram utilizadas o LBP e SVM (LS)

Inicialmente todas as imagens e recortes de imagens do algoritmo proposto foram pré-processadas com o filtro de textura LBP. A técnica utilizada nesta etapa para o algoritmo HA



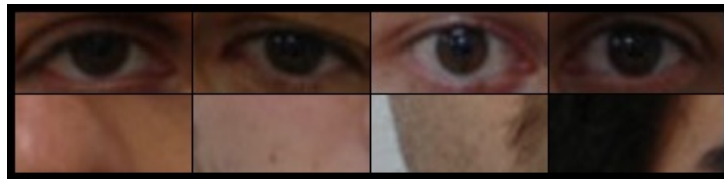
foi a HaarLike Features.

O algoritmo de reconhecimento de padrões utilizado nesta metodologia foi o SVM linear. Para o treinamento foram utilizadas 483 imagens de regiões de olhos e mais 966 de não-regiões de olhos. Na primeira linha da Figura 24 estão exemplos de regiões de olhos e na segunda linha, contra-exemplos. Depois disto, em outro treinamento, foram utilizadas 966 imagens contendo somente os olhos, como os da primeira linha da Figura 25, e o dobro de imagens de não-olho, como na segunda linha da imagem. Desta forma o algoritmo pode realizar as duas etapas da detecção de forma independente. Estas imagens foram recortadas, automaticamente, das imagens que foram previamente normalizadas de forma semiautomática.



**Figura 24: Exemplos de regiões e não-regiões de olhos para treinamento do SVM.**

**Fonte: Autoria Própria.**



**Figura 25: Exemplos de olhos e não-olhos para treinamento do SVM.**

**Fonte: Autoria Própria.**

Em seguida, foram realizadas validações do treinamento ainda utilizando o algoritmo SVM. Para a validação foi utilizado o mesmo número de imagens do treinamento, porém, sem repetir imagens entre os conjuntos. Esta etapa tem como objetivo verificar como o algoritmo se comporta com um conjunto de imagens diferente ao qual ele foi treinado, validando se os parâmetros do SVM são suficientes para classificar as imagens.

Com o algoritmo validado e funcionando, o SVM foi re-treinado juntando o conjunto de treinamento e o conjunto de validação. Ou seja, foram utilizadas 966 imagens para re-treino dos recortes da região dos olhos e 1932 imagens para recortes dos olhos, adicionando o dobro de contra-exemplo para cada um dos treinamentos. O principal objetivo desta etapa é fortalecer o treinamento, aumentando a quantidade de exemplos e contra-exemplos, melhorando a confiabilidade do algoritmo. O SVM treinado foi armazenado em arquivos para serem utilizados nos testes, sem a necessidade de novos treinamentos a cada execução da detecção dos olhos.

O SVM tem como retorno um valor contínuo entre 0 e 1, sendo que, quanto mais próximo do valor 1, maior é a certeza que se tem da imagem apresentada se tratar de um recorte com caso positivo. Em casos extremo, 1 significa certeza de que a região desejada foi detectada, e 0 de que não foi detectada.

Como a proposta é encontrar os olhos numa face humana, foi utilizado um algoritmo de busca, o VNS (Seção 2.3). A busca é feita procurando por uma região ótima na imagem facial que melhor represente uma região de olhos. Inicia-se o processo recortando uma imagem retangular da imagem facial e esta é pré-processada usando o filtro de textura LBP. Assim, o vetor de características da imagem é construído. Este vetor servirá como o exemplo de teste para o SVM que classificará a região cortada como uma região de olhos ou não. Assim, o processo de busca por uma região ótima é repetida através de várias iterações pré-determinadas. A partir do momento em que se tem a região dos olhos, todo o processo de busca é repetido. Agora o espaço de busca passa a ser somente a região encontrada, contendo os dois olhos. Então o VNS procura por duas regiões ótimas que melhor representem os olhos. Sendo assim, as coordenadas centrais de cada um dos olhos são obtidas.

### 3.3 NORMALIZAÇÃO FACIAL

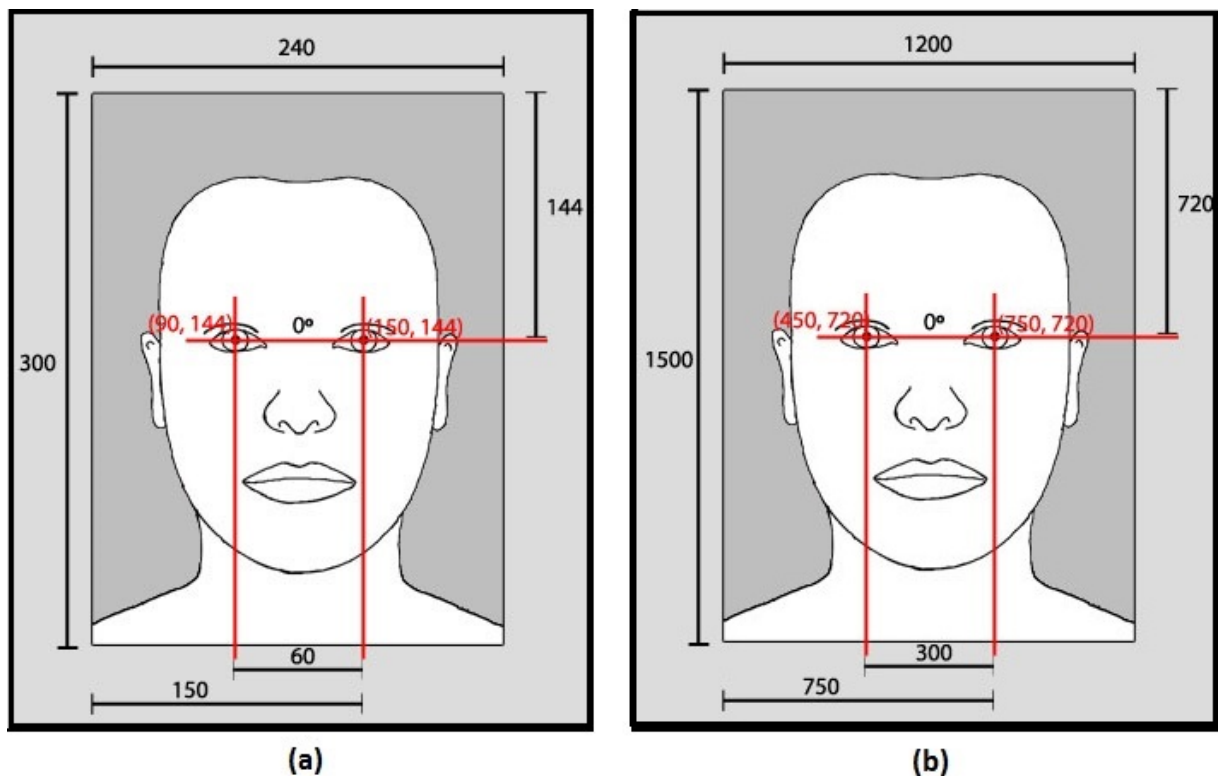
Esta seção apresenta os procedimentos utilizados para a normalização das imagens capturadas. O algoritmo de normalização foi implementado em C/C++ utilizando a biblioteca OpenCV<sup>1</sup>.

As imagens capturadas para a formação da base de imagens possuem 2592x1944 pixels, por isto, e também em função da alta tecnologia das câmeras, a resolução final das imagens normalizadas é maior que o especificado pelas normas (Seção 2.6.1). As imagens são maiores por indicação das especificações, informando que as imagens devem ter no mínimo 240x300, ou seja, podem ser maiores. Com isso o tamanho das imagens normalizadas neste trabalho é de 1200x1500 pixels, ou seja, 5 vezes maior que as indicadas pela norma (Figura 26).

As demais medidas devem seguir a proporção de aumento para serem calculadas, uma imagem exemplo com todos os valores de medidas está disponível na Figura 26 (b). Onde (450, 720) e (750, 720) representam as coordenadas que os olhos devem ter na imagem normalizada e a distância entre eles deve ser de 300 pixels.

---

<sup>1</sup>Open Source Computer Vision. Disponível em: <http://opencv.willowgarage.com/wiki/>.



**Figura 26: Medidas da Imagem Normalizada.**

**Fonte: Autoria Própria.**

### 3.3.1 ETAPAS DA NORMALIZAÇÃO

As etapas da normalização descritas no decorrer desta seção são válidas tanto para semi-automática quanto para a automática. Além disto, elas também são utilizadas para a normalização de faces laterais, diferenciando pelos pontos de referência e dos ângulos adotados durante a normalização.

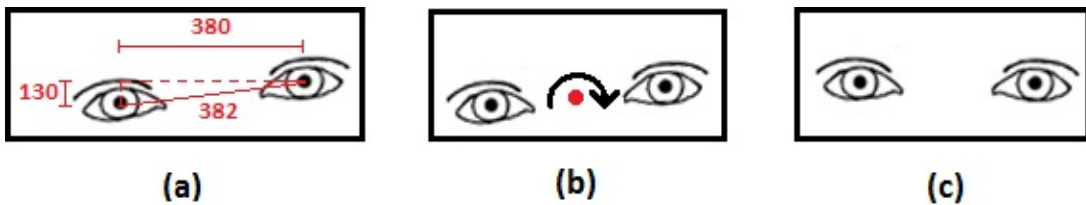
#### 3.3.1.1 ROTAÇÃO

Esta etapa consiste em alinhar os dois olhos da face. Para isto, foi calculada a diferença de pixels entre os dois olhos. Primeiramente, a diferença absoluta entre os valores  $x$  dos pontos e em seguida os valores  $y$ , representando os catetos de um triângulo. Com as medidas dos catetos é possível calcular a hipotenusa (Fórmula 3). Com estes valores é possível obter o ângulo entre os olhos através da Equação 4.

$$\text{hipotenusa} = \sqrt{(\text{cateto oposto})^2 + (\text{cateto adjacente})^2} \quad (3)$$

$$\alpha = \text{acos}\left(\frac{\text{cateto adjacente}}{\text{hipotenusa}}\right) \quad (4)$$

Com isto, basta rotacionar a imagem de acordo com o ângulo obtido, como mostrado na Figura 27. Na Figura 27 (a) os olhos estão posicionados nos pontos (1110, 800) e (1490, 670). Assim, um dos catetos tem o valor 380 e o outro 130, resultando no valor 401 para a hipotenusa. Desta forma, calcula-se o ângulo com a Equação 4, resultando em 18° como resultado. Assim, pode-se rotacionar a imagem a partir do centro dos olhos (Figura 27 (b)), resultando na Figura 27 (c) rotacionada com os olhos nas posições (1100, 740) e (1500, 740), ou seja, alinhados na horizontal.



**Figura 27: Processo de Rotação dos Olhos.**

**Fonte: Autoria Própria.**

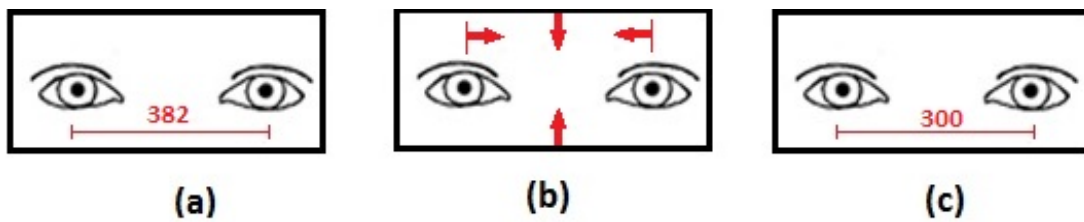
### 3.3.1.2 REDIMENSIONAMENTO

Para que existam 300 pixels entre os olhos é necessário redimensionar a imagem. Neste passo é feita a divisão da distância entre os olhos por 300. Ou seja, basta dividir o valor da hipotenusa, anteriormente calculado, por 300. Com isto, obtém-se o valor de escala necessário para aplicar o zoom na imagem.

A Figura 28 ilustra o processo de redimensionamento. Neste caso a Figura 28 (a) tinha 382 pixels entre os olhos, então o zoom a ser aplicado (Figura 28 (b)) é de aproximadamente 0,78 (300/382). Com isto, resultando numa imagem reduzida com 300 pixels entre os olhos (Figura 28 (c)).

### 3.3.1.3 POSICIONAMENTO DOS OLHOS E RECORTE DA IMAGEM

Através do valor da escala, calculado na etapa anterior, é possível estimar onde os pontos de interesse estão localizados na nova imagem. Assim, é possível calcular o ponto central dos olhos para centralizar a face na imagem. Em seguida, basta cortar a imagem em volta da face, mantendo-a centralizada e obedecendo às medidas citadas anteriormente. A imagem final,



**Figura 28: Processo de Redimensionamento da Face.**

**Fonte: Autorial Própria.**

normalizada, deve ter o tamanho de 1200 X 1500 pixels. O recorte já garante que os olhos estejam nos pontos corretos, (450, 720) e (750, 720).

#### 3.3.1.4 PLANO DE FUNDO

Foi realizado um estudo para normalização de fundo para casos em que o plano de fundo padrão (18% cinza) não abrangesse toda a imagem por trás da face. Estes casos ocorrem na etapa de posicionamento das imagens quando é necessário complementar o fundo para aumentar a imagem. Mesmo colorindo esta parte extra da imagem com a cor cinza, continua havendo uma diferença com o restante. Para isto e para remover demais ruídos foi feito um protótipo de normalização de plano de fundo.

A normalização de plano de fundo foi baseada em operações de segmentação de imagens e morfologia matemática, seguindo a sequência de passos a seguir:

1. Primeiramente a imagem é convertida do formato RGB para o HSV:

Somente a componente H é utilizada no processo, pois é nela que se encontram as principais informações de cor. Esta componente é transformada em uma imagem binária. Nesta etapa a imagem já está segmentada e é possível observar a diferença entre a face da pessoa o fundo da imagem, porém com alguns ruídos;

2. A partir da imagem binária, são utilizadas operações de morfologia matemática:

Primeiramente é aplicado o processo morfológico de Fechamento. Esta etapa tem como objetivo remover os ruídos presentes no plano de fundo. Porém, a região da face também é afetada por esta operação;

Para tentar amenizar o efeito do Fechamento na face, o processo de Abertura é aplicado. Como é o dual do Fechamento, esta etapa tende a voltar a imagem como estava antes, retornando a face ao formato original. Porém, como os ruídos foram totalmente

eliminados na fase anterior, a região compreendida pelo fundo da imagem não sofre alterações;

3. Mesmo com a aplicação da morfologia matemática, ainda existem ruídos dentro da face. Para contornar este problema uma nova segmentação é feita, agora utilizando o *Region Growing*. Dois pontos são definidos nas extremidades superiores da imagem, onde há a certeza que existe plano de fundo. Então, o *Region Growing* é aplicado, sendo que ele para de agir na imagem quando encontra a região da face, pois não existe mais conectividade entre os pixels. Ou seja, a parte interna da face fica totalmente separada do fundo e sem nenhum ruído;
4. Com isto resulta numa imagem segmentada em duas regiões, onde uma delas compreende a face. Então, a região compreendida pela face é substituída pela imagem original, resultando somente na face com um fundo normalizado contendo somente uma cor.

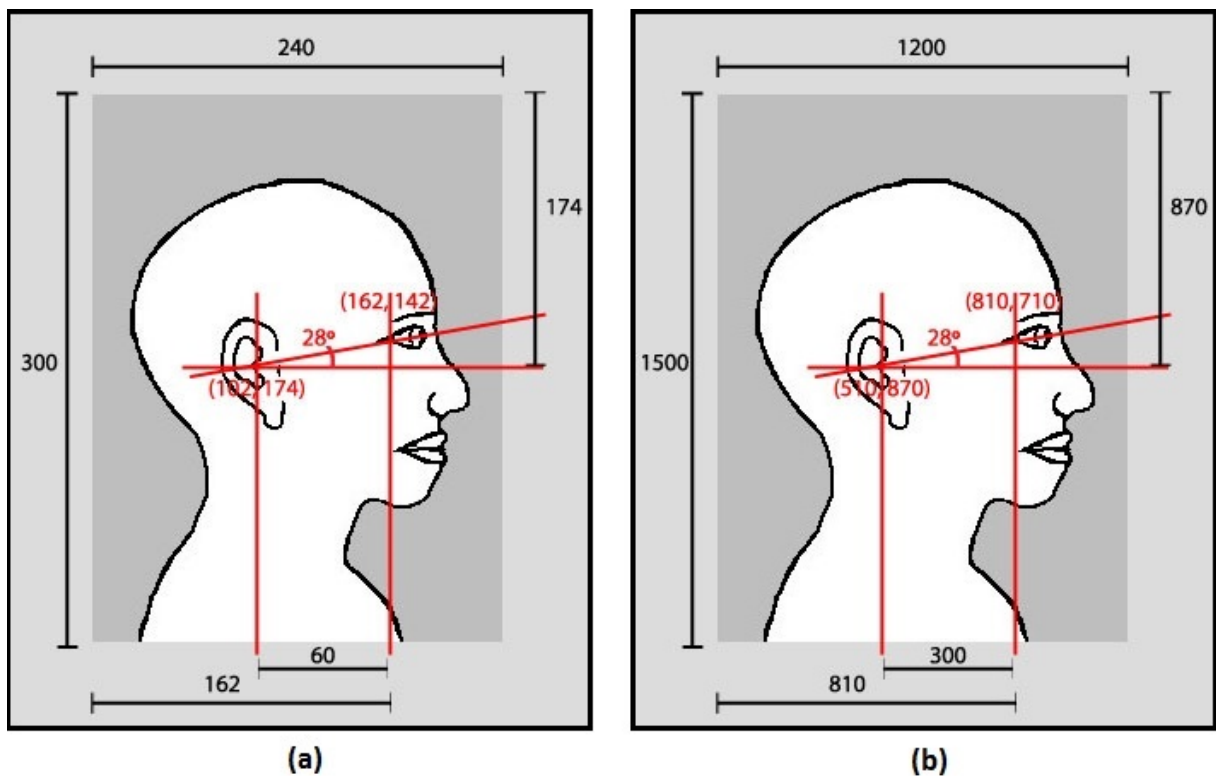
### 3.3.2 NORMALIZAÇÃO DE FACES LATERAIS

Além da normalização frontal das imagens que foi caracterizada na seção anterior, é proposto um novo procedimento de normalização das faces laterais, em perfil. Seguindo os procedimentos estabelecidos pelas regras da normalização facial frontal, buscou-se adaptar estas regras visando a normalização lateral. As imagens de face em perfil podem ser utilizadas como fonte de características diferentes da face frontal, ou ainda, as duas imagens podem ser utilizadas em conjunto a fim de obter uma maior precisão no RF. Os procedimentos utilizados são explicados no decorrer desta seção.

A ideia geral adotada para a nova metodologia de normalização de faces laterais foi a mesma utilizada anteriormente para a normalização de faces frontais. Primeiramente foram estabelecidos dois pontos de referência para realizar o processo de normalização. Em seguida, utilizando os pontos selecionados, deve-se rotacionar, redimensionar e posicionar a face na imagem e, por fim, recortar a imagem para obter os tamanhos desejados (Figura 29).

A Figura 29 (a) mostra todas as medidas, ângulos e pontos de referência da imagem normalizada lateral canônica. Assim como na imagem normalizada frontal canônica, estes valores podem ser aumentados proporcionalmente, a fim de obter uma imagem normalizada com maior resolução. O tamanho da imagem canônica deve ser de 240x300 e as demais medidas da Figura 29 estão explicadas a seguir.

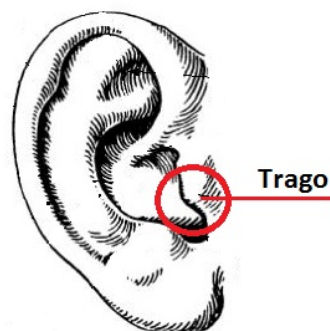
Primeiramente, foram escolhidos os dois pontos de referência para iniciar o processo de normalização. Como a face está em perfil, não é possível obter pontos dos dois olhos da



**Figura 29: Medidas da Face Lateral Normalizada.**

**Fonte: Autorial Própria.**

face, por isso, somente um dos olhos será utilizado como referência e o outro ponto escolhido localiza-se na região da orelha. O ponto mais fácil de encontrar nas orelhas, de modo geral, é o trago (Figura 30), sendo assim, o canto do olho e o trago foram utilizados como pontos de referência.



**Figura 30: Trago - Região da Orelha.**

**Fonte: Autorial Própria.**

A partir dos dois pontos de referência, a face pode ser rotacionada. No caso da normalização lateral, os pontos de referência não ficarão alinhados como no caso das faces

frontais, pois quando eles estão alinhados a face fica voltada para baixo. Então foi adotado um ângulo médio para as imagens da base com o objetivo de manter a face reta, como se o indivíduo estivesse olhando para frente, sem inclinações da cabeça. O ângulo satisfatório foi fixado empiricamente em  $28^\circ$  de acordo com estudos na base de imagens.

A etapa de redimensionamento consiste em padronizar o tamanho da imagem. Nas imagens laterais, a distância horizontal entre os dois pontos de referência deve ser de 60 pixels, como se observa na Figura 29 (a).

Em seguida, a imagem é recortada de modo que fique com 240x300 pixels e os pontos de referência fiquem posicionados no lugar correto. Sendo que, para as imagens laterais direitas o trago deve ficar no ponto (102, 174) e o canto do olho no ponto (162, 142) (Figura 29 (a)).

As imagens laterais também foram normalizadas num tamanho cinco vezes maior do que a imagem canônica, assim como nas imagens frontais, isto pode ser observado na Figura 29 (b). Com isto, o tamanho final das imagens é de 1200x1500 pixels, com uma distância horizontal de 300 pixels entre os pontos de referência. Onde o trago ficou localizado no ponto (510, 870) e o canto do olho no ponto(810, 710).

### 3.3.3 NORMALIZAÇÃO SEMIAUTOMÁTICA

A normalização semiautomática teve como etapa adicional a obtenção dos pontos centrais dos olhos, além da normalização propriamente dita. Os pontos centrais são utilizados para comparação com os olhos encontrados na etapa de detecção de olhos (Seção 3.2). Assim, é possível estabelecer a diferença entre o ponto central obtido manualmente e o obtido pela detecção de olhos, sendo possível comparar a precisão dos algoritmos. Além disto a normalização semiautomática pode ser utilizados para casos da normalização automática em que a detecção de ROI não encontre os pontos desejados.

Esta etapa é considerada semiautomática pois os pontos centrais de cada olho foram obtidos de forma manual. Foi desenvolvido um sistema, em C++ com OpenCV, que apresenta a imagem na tela para que sejam selecionados os pontos de interesse. Em seguida, as coordenadas dos olhos são salvas em um arquivo e o processo de normalização é executado automaticamente.

No caso da normalização semiautomática frontal foi optado por selecionar 4 pontos para maior precisão. Para ambos os olhos foram selecionados um ponto à esquerda e um ponto à direita da íris, como observado nos pontos pretos da face da Figura 31. Estes pontos foram selecionados pensando em uma linha que os une, passando pela pupila (linha branca na Figura 31). E então, o ponto médio é considerado o ponto central do olho, ponto em branco na Figura



31. O motivo principal para escolher 2 pontos por olho é que em alguns casos o olho pode estar um pouco fechado e a íris pode ser escura. Isto dificulta encontrar a localização de um ponto central ou da pupila, como em alguns casos de pessoas orientais, como foi possível observar durante os experimentos. Adotando a metodologia de selecionar 2 pontos por olhos foi possível contornar o problema.



**Figura 31: Seleção dos pontos de referência para normalização frontal.**

**Fonte: Autoria Própria.**

As imagens laterais também foram normalizadas de forma semiautomática seguindo a mesma ideia da normalização frontal. Para seleção das referências são escolhidos apenas dois pontos, que já são os pontos utilizados na normalização. Isto se deve ao fato de ser mais fácil localizar o canto do olho para selecionar um dos pontos, mesmo nos casos em que o olho está um pouco fechado ou no caso dos olhos de pessoas orientais. O local dos pontos pode ser observado na face da Figura 32, destacados em preto.



**Figura 32: Seleção dos pontos de referência para normalização lateral.**

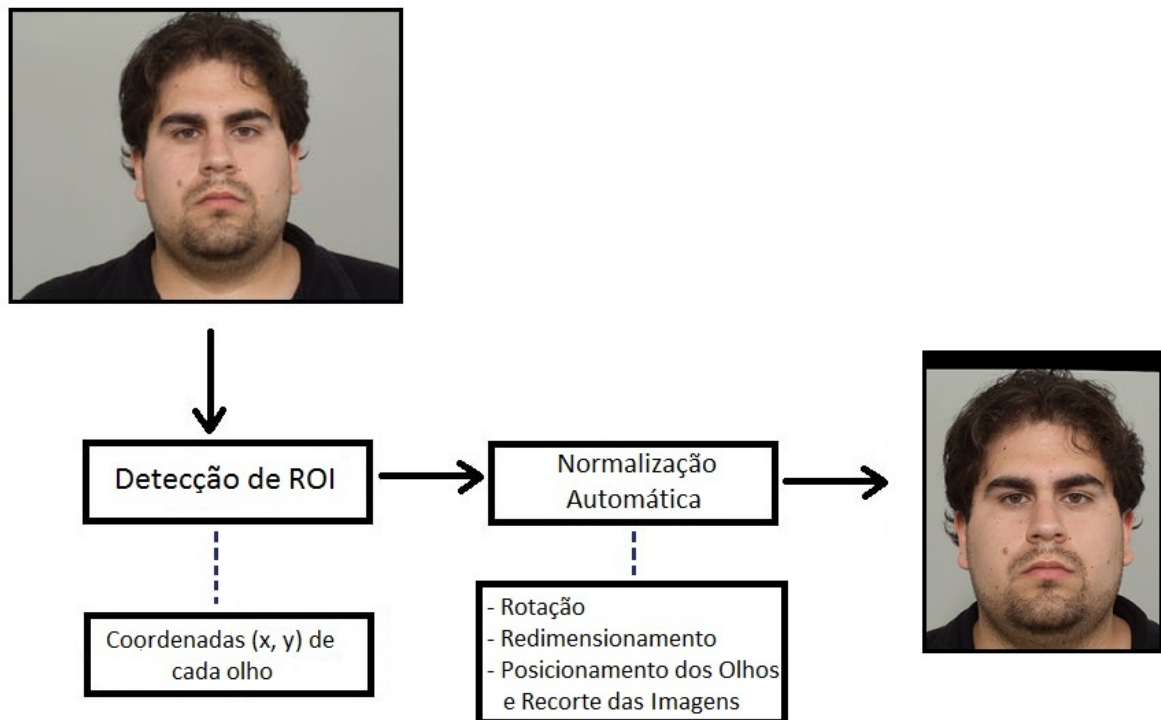
**Fonte: Autoria Própria.**

### 3.3.4 NORMALIZAÇÃO AUTOMÁTICA

O processo de normalização automática utiliza duas etapas anteriormente vistas, as etapas de normalização de imagens (Seção 3.3.1) e a detecção de ROI (Seção 3.2). Esta nova

metodologia é diferente da semiautomática pelo fato dos olhos serem detectados de forma automática e não manual. A normalização automática facial foi aplicada somente para as imagens frontais.

Esta metodologia funciona como apresentado na Figura 33. Utilizando todos os conceitos já apresentados de detecção e normalização, a imagem é apresentada ao sistema e este de forma automática identifica as ROIs. Com isto, os dois pontos de interesse (olhos) são apresentados para a normalização, agindo de forma automática a fim de padronizar a imagem.



**Figura 33: Processo automático de normalização.**

**Fonte: Autorial Própria.**

## 4 EXPERIMENTOS E RESULTADOS

Os resultados obtidos neste estudo são apresentados nas seções seguintes, divididos nas etapas principais deste trabalho. Na primeira seção (Seção 4.1) são apresentados os resultados da Construção da Base de Imagens Faciais; na Seção 4.2 são mostrados os resultados da Detecção de Olhos e, em seguida, os resultados sobre a Normalização das Imagens (Seção 4.3).

### 4.1 BASE DE IMAGENS FACIAIS

O processo de criação da base de imagens faciais foi realizado seguindo as instruções descritas anteriormente (Seção 3.1). Todo o processo de Construção da Base de Imagens Faciais levou aproximadamente nove meses para ser concluído. Neste período estão incluídos a escolha do local e adaptação do ambiente, a busca por voluntários para participar do processo, os testes iniciais antes de começar a criação da base definitiva, a captura das imagens e a seleção das imagens que foram escolhidas para compor a base de imagens faciais final.

Testes e validações de captura de imagens foram feitos com um número reduzido de voluntários a fim de verificar a fidedignidade do processo e a qualidade das imagens que formariam a base de imagens. Alguns exemplos destas sequências de testes podem ser observados nas imagens da Figura 34, sendo que:

1. Na primeira sequência de testes (primeira linha de imagens da Figura 34), as imagens apresentam sombra no plano de fundo. E as imagens laterais ainda estavam sem definição de como seriam capturadas, onde somente o rosto estava de lado, sem rotacionar todo o corpo. Neste teste foram capturadas imagens de 7 voluntários, totalizando 42 imagens;
2. Na segunda sequência de testes (segunda linha), a iluminação se mostrou excessiva e a sombra aparece com mais evidência no fundo. A partir deste teste, já foram definidos os procedimentos para captura das fotos laterais, com o corpo virado junto com a face. Isto ocorre porque quando se rotaciona somente a face existe a possibilidade de não ficar totalmente na lateral, ou seja, pode ficar com menos que 90° e ainda não seria possível



**Figura 34: Exemplos de seqüências de imagens dos testes preliminares.**

**Fonte: Autorial Própria.**

capturar as imagens com  $105^\circ$  para a maioria das pessoas. Neste teste foram capturadas imagens de 5 pessoas, num total de 70 imagens;

3. Na última seqüência de teste antes da base definitiva, terceira linha da Figura 34, a iluminação ficou fraca, resultando em menor incidência de sombra no plano de fundo. Este teste foi feito somente com 2 voluntários, num total de 52 imagens.

A partir disto, foram finalmente definidos os detalhes para a captura definitiva das imagens para a criação da base, como foi apresentado na metodologia da Seção 3.1. Os resultados com os detalhes finais são apresentados no decorrer desta seção, incluindo as imagens que não foram satisfatórias para a base de imagens faciais final.

O ambiente escolhido e já pronto com os equipamentos de acordo com as normas, é isolado de qualquer fonte de iluminação externa. Os equipamentos foram dispostos como foi explicado anteriormente na Figura 20. O ambiente final, com os equipamentos devidamente dispostos, pode ser observado na Figura 35.

Apesar do cuidado tomado durante a captura, algumas imagens não ficaram como exigido pelas normas, então foram removidas da base, sendo que para alguns voluntários elas foram re-capturadas para compor a base. Alguns exemplos são mostrados nas imagens da Figura 36 e os principais motivos que levaram à remoção das imagens da base foram:

1. plano de fundo e/ou a própria imagem com coloração alterada (Figura 36 (a)), pois a luz era refletida por causa cor da roupa, por exemplo;



**Figura 35: Ambiente da captura de imagens.**

**Fonte: Autoria Própria**

2. presença de barba cobrindo grande parte da face (Figura 36 (b));
3. cabelo escondendo parte da orelha (Figura 36 (c));
4. cabelo cobrindo parte da sobrancelha e da testa (Figura 36 (d));



**Figura 36: Exemplos de imagens excluídas da base final.**

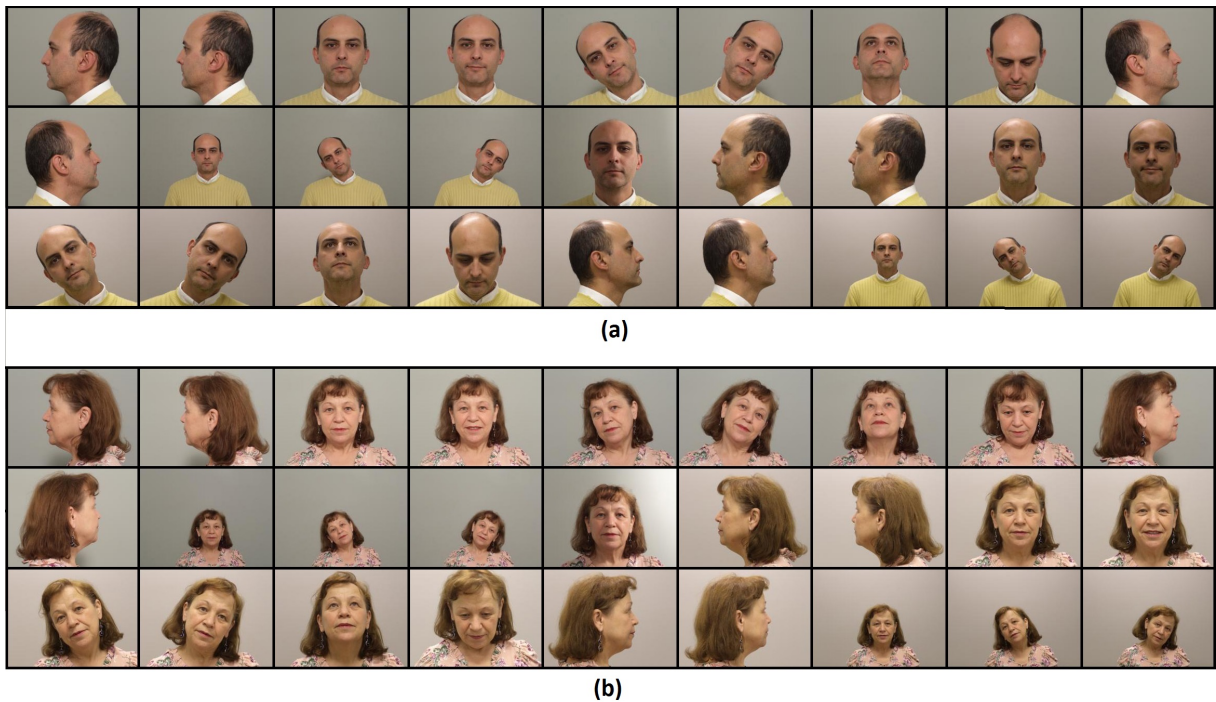
**Fonte: Autoria Própria.**

5. olhos fechados (Figura 36 (e));
6. voluntário olhando para os lados e não para a câmera (Figura 36 (f)).

Devido ao fato de existirem algumas imagens com os problemas citados, foi realizada uma validação manual com o objetivo de padronizar a base de acordo com as normas estabelecidas. A partir disto, as imagens indesejadas foram separadas e então a base de imagens faciais foi construída. As imagens da Figura 37 representam exemplos de imagens que compõem a base definitiva. Nesta Figura estão presentes todas as imagens de dois voluntários, na sequência apresentada na Tabela 4. Na Figura 37 (a) as 27 imagens para um voluntário do sexo masculino e na Figura 37 (b) do sexo feminino.

De modo a organizar a base, todas as imagens receberam uma codificação (Seção 3.1.4.1). Um exemplo da codificação é a imagem com o nome '007-1-12-1-3-0.jpg', cada código é explicado na Tabela 7, conforme seus significados vistos na Tabela 5.

Finalmente, podemos observar na Tabela 8 um resumo sobre as características da Base



**Figura 37: Exemplos de imagens da base facial construída.**

**Fonte: Autoria Própria.**

**Tabela 7: Exemplo Codificação dos Arquivos de Imagem.**

Código	Informação
007	código que representa a pessoa, visando preservar seu nome
1	indica que é o cenário C1
12	*quando a imagem for 14 o cenário é C2, mas o código de iluminação ainda é 1 representa que a pose da pessoa é com inclinação para direita, sendo que a câmera está afastada 2,5m da pessoa
1	indica que o voluntário é do sexo masculino
3	faixa etária da pessoa é entre 31 e 40 anos (Tabela 6)
0	código auxiliar, até o momento todas as imagens possuem o mesmo valor
.jpg	extensão de arquivos adotada para salvar as imagens

**Fonte: Autoria Própria**

de Imagens Faciais construída neste trabalho. No total foram capturadas 5775 fotos, onde, 2049 imagens foram removidas da base oficial, sendo que esta possui 3726 imagens (Tabela 8). As imagens excluídas da base ainda estão armazenadas para eventuais testes de algoritmos. O número total de participantes foi de 191, mas na base final aparecem 138, onde 75 são homens e 63 mulheres. As demais características da base, resolução, imagens por pessoa, poses e iluminações, já foram apresentadas anteriormente.

**Tabela 8: Resumo das características da Base de Imagens Faciais Construída.**

Resolução	2592x1944
Pessoas	138
Imagens por Pessoa	27
Poses	13
Iluminações	3
Total de Imagens	3726
Homens	75
Mulheres	63

**Fonte: Autoria Própria**

#### 4.2 DETECÇÃO DE OLHOS

Os experimentos foram realizados utilizando a base de imagens faciais desenvolvida neste trabalho. Para os testes de detecção foram utilizadas as imagens frontais divididas em seis sequências separadas por pose (Tabela 4), com 138 imagens cada, como seguem:

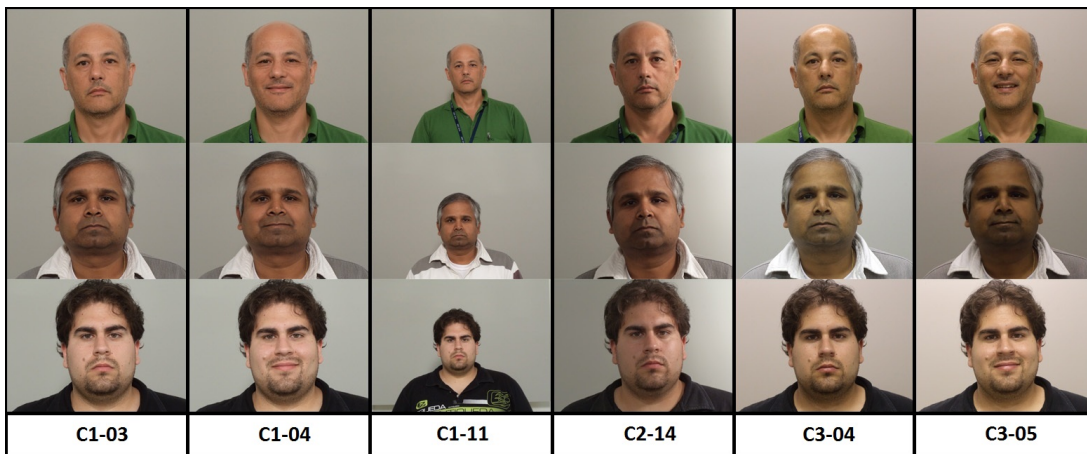
- C1-03: cenário C1 e imagens frontais;
- C1-04: para frontais com expressão do C1;
- C1-11: C1 e imagens frontais em escala;
- C2-14: cenário C2;
- C3-03: imagens frontais sem expressão do cenário C3;
- C3-04: C3 e frontais com expressão.

A Figura 38 apresenta três exemplos das sequências de testes utilizadas. As Tabelas 9 e 10 apresentam os resultados obtidos nos experimentos.

No primeiro algoritmo, HA, as soluções não variam quando o algoritmo é executado mais de uma vez. Já no segundo algoritmo as soluções mudam, pelo fato de se utilizar uma técnica de busca onde existe uma certa aleatoriedade para iniciar as variáveis de busca. Por isto, na Tabela 10, os valores mostrados são o resultado da média entre várias execuções. O desvio padrão da média das execuções é mostrado no gráfico da Figura 39. Com este gráfico é possível comparar as diferenças de acertos entre os algoritmos.

Com os resultados apresentados (Tabela 9), é possível notar que a metodologia HA é eficiente tanto no número de acertos quanto no tempo de processamento para encontrar os





**Figura 38:** Exemplos das imagens utilizadas nos experimentos de detecção de olhos.

**Fonte:** Autoria Própria

**Tabela 9:** Resultados da Detecção de Olhos com o algoritmo HA.

Pose	Tempo Médio/imagem (s)	Tempo Total (s)	Acertos	Porcentagem de acerto
C1-03	2,2	305	131	94,9%
C1-04	2,7	373	126	91,3%
C1-11	2,1	300	136	98,5%
C2-14	2,7	385	134	97,1%
C3-03	2,9	404	137	99,2%
C3-04	2,2	304	135	97,8%

**Fonte:** Autoria Própria.

olhos. O maior número de acertos foi para a sequência C3-03, em compensação foi o que mais demorou, em média, para encontrar os olhos na face. No geral, o melhor resultado pode ser observado com o teste C1-11, onde o tempo médio foi o menor dentre os testes, apresentando erros somente em duas faces.

**Tabela 10:** Resultados da Detecção de Olhos com o algoritmo LS.

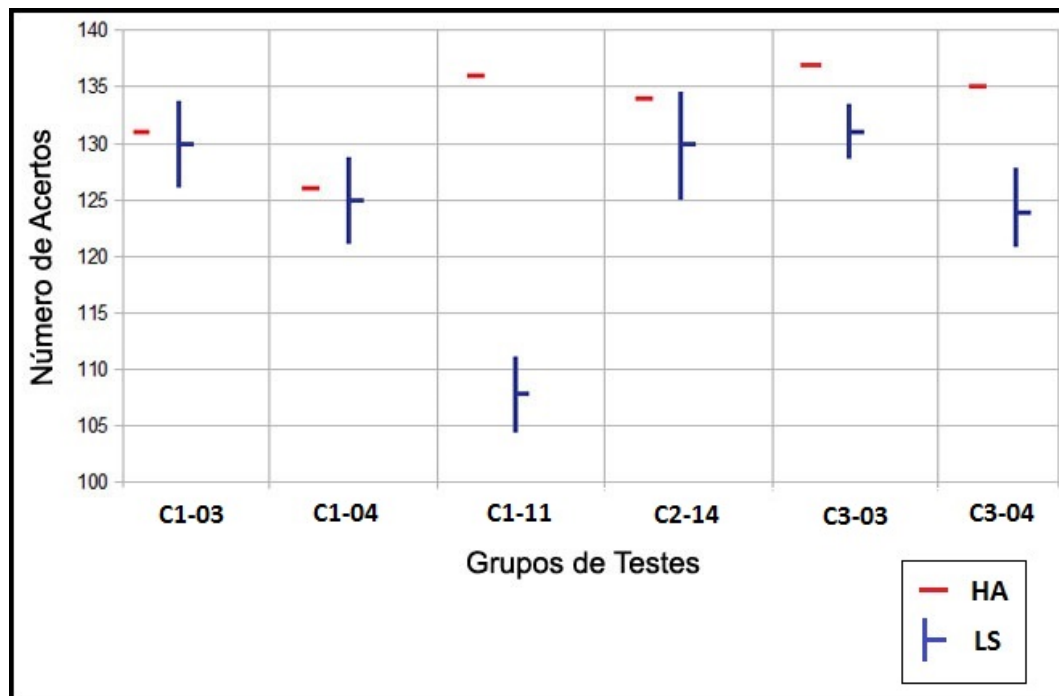
Pose	Tempo Médio/imagem (s)	Tempo Total (s)	Acertos	Porcentagem de acerto
C1-03	26,5	3664	130	94,2%
C1-04	26,9	3725	125	90,5%
C1-11	26,9	3721	108	78,2%
C2-14	22,7	3135	130	94,2%
C3-03	27,1	3753	131	94,9%
C3-04	27,5	3795	124	89,9%

**Fonte:** Autoria Própria.

A Tabela 10 mostra as médias de 20 rodadas de execução do algoritmo LS. A partir dos dados compilados na tabela, observa-se que este algoritmo é mais lento. Porém, na maioria

dos casos é eficiente quanto ao número de acertos. Em média, o maior número de acertos foi 131 (C3-03) com tempo médio por imagem de aproximadamente 27 segundos para encontrar os olhos. Porém, nas imagens com escala (sequência de teste C1-11), o número de acertos foi 108, um valor reduzido quando comparado ao algoritmo HA que obteve 136 acertos.

O gráfico da Figura 39 mostra os resultados de todos os testes, onde cada coluna representa uma categoria de imagens de teste. Sendo que os teste do algoritmo HA sempre apresenta o mesmo número de acertos, independente de número de execuções. Isso está representado onde existem somente linhas horizontais no gráfico. Como na segunda abordagem é utilizado o VNS os valores variam de acordo com as execuções, portanto, os resultados são mostrados usando o desvio padrão (linha vertical) e a média (linha horizontal) dos testes.



**Figura 39: Gráfico de comparação entre os acertos dos algoritmos.**

**Fonte: Autoria Própria.**

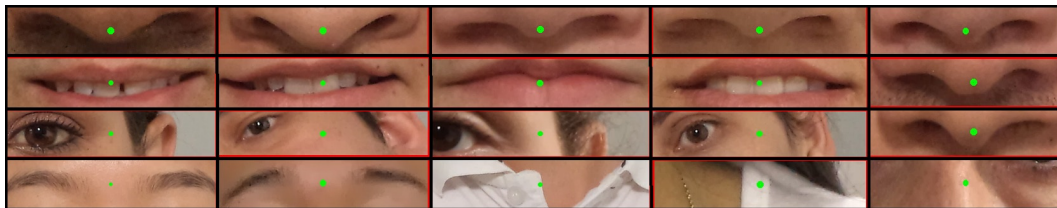
Com o gráfico (Figura 39) é possível notar que em relação às médias de acertos o algoritmo HA foi o melhor. Levando em consideração o desvio padrão, o algoritmo LS pode ser comparado com o HA e foi melhor em alguns resultados:

- No teste C1-03, onde o desvio padrão foi de 3, no melhor caso o número de acertos seria 133 contra 131 do HA;
- Já no teste C1-04, com o desvio padrão também de 3, o algoritmo LS acertou, no melhor caso, 128 contra 125 do HA;

- Para o teste C2-14 foram 135 acertos no melhor caso para o algoritmo LS, sendo que o desvio padrão teve valor 5, e 134 acertos no HA;
- Para os demais testes, mesmo considerando o desvio padrão, a metodologia HA foi o melhor em quantidade de acertos.

Comparando os algoritmos no que se refere ao tempo de processamento, pode-se afirmar que o algoritmo HA é mais eficiente. De modo geral, este algoritmo foi aproximadamente 10 vezes mais rápido que o algoritmo proposto. Isto se deve ao fato da aleatoriedade presente na busca do VNS em conjunto com as diversas validações que devem ser feitas nos recortes das imagens através do algoritmo SVM.

Alguns exemplos de regiões que foram detectadas erroneamente pelos algoritmos podem ser observadas na Figura 40. A região do nariz foi a que induziu maior erro de detecção dentre todas as regiões detectadas de maneira equivocada. Isto se deve ao fato de esta região ter uma textura semelhante com a região dos olhos, sendo percebido como dois círculos escuros no meio de uma região clara.



**Figura 40: Regiões detectadas erroneamente.**

**Fonte: Autoria Própria.**

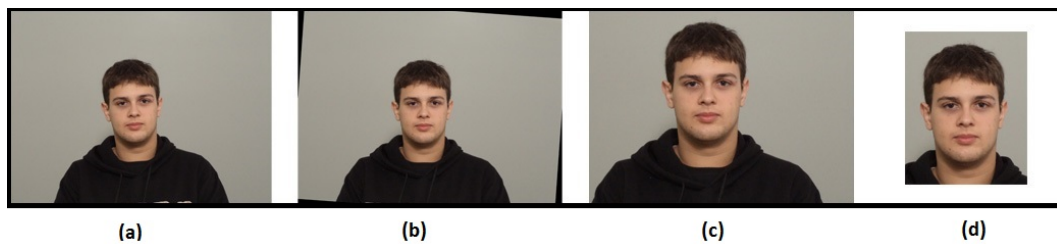
### 4.3 NORMALIZAÇÃO FACIAL

A normalização facial foi realizada seguindo as etapas apresentadas na Seção 3.3.1. A normalização frontal foi realizada primeiramente de forma semiautomática, o principal objetivo era obter as coordenadas dos olhos para utilizar na validação dos algoritmos de detecção de ROI. Em seguida foram realizados experimentos juntando a detecção das RoIs com a normalização, gerando a normalização automática. Neste processo, quando os olhos não são encontrados, a normalização semiautomática pode ser utilizada novamente. Já a normalização lateral foi feita somente de forma semiautomática. Foram normalizadas as 7 imagens frontais e as 2 laterais direita de cada indivíduo. Alguns resultados das normalizações são mostrados a seguir.

### 4.3.1 NORMALIZAÇÃO SEMIAUTOMÁTICA FRONTAL

Um exemplo do resultado do passo a passo da normalização aplicada neste trabalho pode ser observado nas Figuras 41 e 42. Na Figura 41 (a) um exemplo de uma imagem que foi rotacionada (Figura 41 (b)) e ampliada (Figuras 41 (c)) para se adequar ao tamanho exigido pelas normas. Em seguida, foi recortada (Figuras 41 (d)) para ficar com as dimensões desejadas. É possível observar que, apesar de ampliar a imagem, a imagem normalizada é menor que a original (a). Neste caso a imagem é ampliada só para adequar o tamanho da face, apesar da imagem ser menor, a face é maior que do a original.

Já no caso da Figura 42 (a) ocorreu a redução da imagem (Figura 42 (c)) para adequar ao tamanho desejado pelas normas (Seção 2.6). Ainda é possível notar que as as imagens normalizadas finais ficaram do mesmo tamanho e com as faces praticamente iguais (Figuras 41 (d) e 42 (d)).



**Figura 41: Etapas da normalização frontal, com ampliação da imagem.**

**Fonte: Autoria Própria.**



**Figura 42: Etapas da normalização frontal, com redução da imagem.**

**Fonte: Autoria Própria.**

Foram normalizadas todas as imagens faciais frontais presentes na base. Com isto, a normalização frontal resultou numa base com 966 imagens, onde cada indivíduo tem 7 imagens normalizadas. Dentre estas imagens estão quatro com iluminação própria (sem expressão, com expressão, com escala e com iluminação lateral) e mais duas com a iluminação ambiente (sem expressão e com expressão). Nesta sequência, algumas imagens com resultados da

normalização podem ser vistas na Figura 43, onde observa-se as imagens frontais normalizadas para dois voluntários da base.



**Figura 43: Exemplos de imagens frontais normalizadas.**

**Fonte: Autorial Própria.**

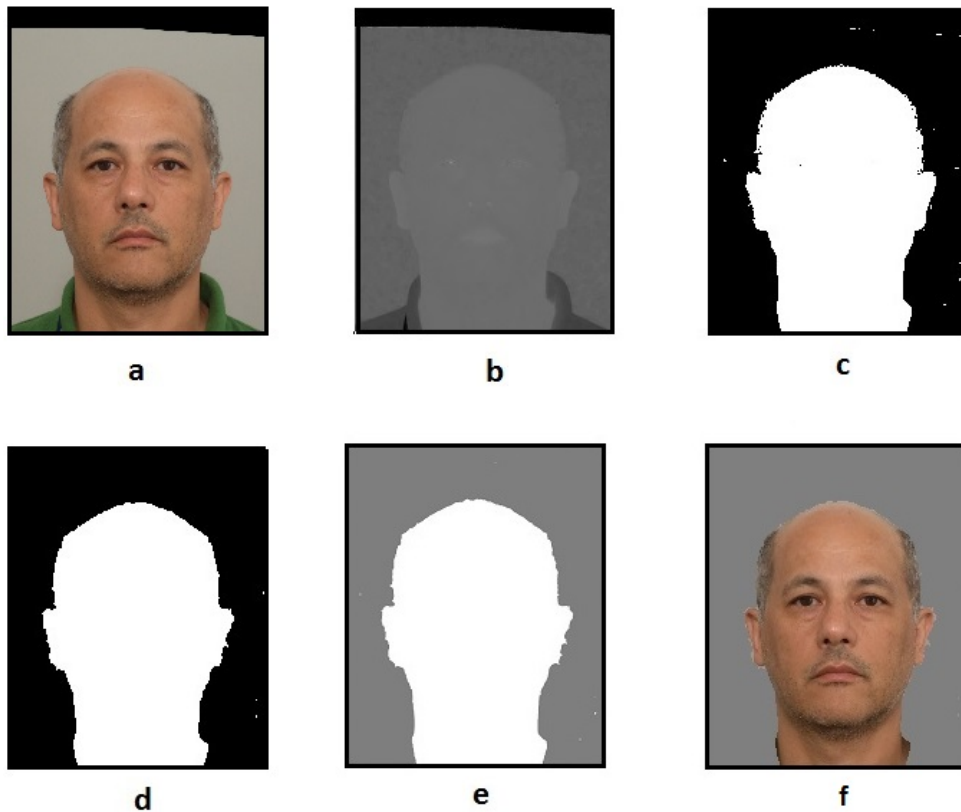
#### 4.3.1.1 NORMALIZAÇÃO DE PLANO DE FUNDO

A Figura 44 apresenta as etapas do Processo de Normalização do Plano de Fundo. A primeira imagem, Figura 44 (a), é a original ainda no formato RGB. A imagem original é transformada em HSV, a Figura 44 (b) representa a componente H e a Figura 44 (c) sua forma binária. O valor do limiar utilizado para binarização é o pixel que fica entre o meio dos olhos, isto era calculado para cada imagem de forma individual.

Pode-se observar que existem vários ruídos na região preta, que compreende o plano de fundo. Para resolver este problema e ao mesmo tempo manter a forma da face, o fechamento e a abertura são aplicados em sequência, resultando na Figura 44 (d) quase sem ruídos. Para as operações morfológicas foram utilizados elementos estruturantes iguais, sendo o Rhombus (Figura 5) escolhido para tentar manter o formato das regiões curvas da face. O tamanho do elemento estruturante era de 5 pixels para cada lado e foram executadas 5 iterações, tanto para o fechamento quanto para a abertura.

Finalmente, é aplicado o *Region Growing* já com o nível de cinza correto indicado pelas normas, o cinza 18%. Com isto, mais alguns ruídos são removidos e a imagem contém duas regiões segmentadas, plano de fundo e face (Figura 44 (e)). Na última etapa a área compreendida pela face desta imagem é substituída pela imagem original, resultando na imagem normalizada da Figura 44 (f).

Este processo de normalização não foi eficiente para a maioria das imagens da base facial. Para algumas imagens, quando o cabelo da pessoa era claro, a componente H não seg-



**Figura 44: Etapas da Normalização de Plano de Fundo.**

**Fonte: Autoria Própria.**

mentava a imagem da forma desejada, incluindo o cabelo no plano de fundo. O mesmo acontecia para algumas imagens com a iluminação lateral, por ter um lado da face muito claro, este era misturado com o fundo. Em outros casos os processos de morfologia matemática acabaram deformando elementos da face. Outro problema na normalização do plano de fundo era quando não estava bem definida a borda entre a face e o fundo, fazendo com que o *Region Growing* encontrasse um pixel conectado com o fundo, incluindo algumas partes da face no plano de fundo. Alguns exemplos de sucesso da normalização são mostrados na primeira linha da Figura 45 e de imagens que não foram normalizadas corretamente na segunda linha.

#### 4.3.2 NORMALIZAÇÃO SEMIAUTOMÁTICA LATERAL

Exemplos do passo a passo da normalização das imagens laterais aplicada neste trabalho podem ser observados nas Figuras 46 e 47. Na etapa de rotação a Figura 46 (b) foi rotacionada para baixo com intuito de obter os  $28^\circ$  entre os pontos de referência. Já na Figura 47 (b) a face foi rotacionada para cima.



**Figura 45: Exemplos da Normalização de Plano de Fundo.**

**Fonte: Autoria Própria.**

Ambas imagens foram reduzidas (Figuras 46 (c) e 47 (c)) para obter o tamanho desejado e em seguida recortadas (Figuras 46 (d) e 47 (d)) para ficar com as dimensões de uma imagem normalizada.



**Figura 46: Etapas da normalização lateral, com rotação para cima.**

**Fonte: Autoria Própria.**



**Figura 47: Etapas da normalização lateral, com rotação para baixo.**

**Fonte: Autoria Própria.**

Os resultados para algumas imagens podem ser vistas na Figura 48. Esta normalização resultou numa base com 276 imagens, sendo 2 por participante. Foram normalizadas somente as imagens laterais direita a critério de estudo, sendo uma com iluminação própria para captura de imagens e outra com iluminação ambiente.



**Figura 48: Exemplos de imagens laterais normalizadas.**

**Fonte: Aatoria Própria.**

#### 4.3.3 AUTOMÁTICA COM DETECÇÃO DE OLHOS

Os testes desta fase são os mesmos que já foram realizados com a normalização semiautomática. Porém, agora a detecção de ROIs fornece as coordenadas para a etapa de normalização. Como existem erros na detecção de ROIs, algumas imagens não puderam ser normalizadas de forma automática. Para estes casos é necessário realizar a seleção dos pontos de interesse de forma manual, de modo que a imagem seja normalizada corretamente. De modo geral a normalização automática é válida, pois a porcentagem de acerto dos algoritmos de detecção de ROIs foi alta e somente para os casos de erro seria necessária a normalização semiautomática.



## 5 CONSIDERAÇÕES FINAIS

Com a execução deste trabalho foi possível chegar aos objetivos desejados, ou seja, foi construída uma nova base de imagens faciais levando em consideração as normas internacionais. E, com isso, foi possível realizar a normalização de imagens da base, gerando uma base com imagens normalizadas. Sendo que para a normalização automática das faces foram implementadas metodologias de detecção de ROI.

### 5.1 DIFICULDADES ENCONTRADAS

A principal dificuldade encontrada na construção da base de imagens faciais foi encontrar voluntários dispostos a terem suas imagens capturadas e publicadas cientificamente. Outro fator que teve uma maior trabalho foi adequar o ambiente de captura, assim como ajustar as configurações da câmera e da iluminação. Selecionar as imagens boas para construção da base e também a organização desta, com uma codificação ideal e de fácil compreensão, foram etapas que demandaram um maior tempo.

Já para a Detecção de ROIs, o principal problema foi em definir um intervalo em que os algoritmos deveriam variar em escala, modificando o tamanho das imagens, de modo a procurar a região de olhos nas imagens menores.

Adequar a normalização do plano de fundo foi um processo que exigiu a combinação de várias técnicas e foi trabalhoso adequar uma sequência de etapas para que o plano de fundo ficasse separado da face.

### 5.2 CONCLUSÕES

A base de imagens faciais resultou em faces com diversas variações, como iluminação, pose, expressão e distância, além das imagens normalizadas frontal e lateral. Com a captura de imagens foi possível explorar outros assuntos quando se trata de Visão Computacional. Este tema abrange diversas etapas, que se estendem desde o entendimento sobre uma configuração

da câmera para obter uma melhor imagem até a organização de um ambiente adequado para capturar imagens em diversas condições, controladas ou não.

O mesmo pôde ser observado durante o processo de normalização, pois mesmo com as imagens capturadas nas melhores condições, é possível adequá-las de acordo com a aplicação em que serão utilizadas.

Apesar de parecerem triviais, as tarefas de captura e normalização exigem cuidados. Tanto que, para ambas tarefas, algumas normas devem ser seguidas e observa-se que muitas das bases existentes não seguem estas normas. O estudo e aplicação das normas resultou em certas adaptações, ou seja, além dos ajustes previstos nas normas foram criadas alternativas que atendessem aos critérios necessários para este estudo.

De modo geral, ambos algoritmos testados foram eficientes para detectar olhos em imagens de faces humanas. Porém o algoritmo HA é mais rápido e possui uma eficiência maior quanto a encontrar de fato os olhos.

Com os resultados foi possível construir a base de imagens faciais, tanto a comum quanto a normalizada. Para este processo foram executadas as três etapas iniciais do processo de RF. São elas: a captura de imagens faciais, a detecção de ROIs e a normalização automática das faces.

### 5.3 TRABALHOS FUTUROS

Uma etapa interessante a continuar sendo estudada é a normalização do plano de fundo. Está etapa pode vir a auxiliar o processo de RF. Outra etapa da normalização a ser estudada é em relação a iluminação, buscando adequar as imagens com problemas de iluminação, como sombras e iluminação fraca ou irregular.

Para melhorar os sistema de normalização automática como um todo, aperfeiçoar e ajustar os algoritmos para detectar as imagens. Com isto melhorando as taxas de acerto, buscando diminuir o uso da normalização semiautomática para quando os olhos não são detectados corretamente.

Em breve, as bases de imagens construídas, com imagens normais e normalizadas, deve ser disponibilizada de forma pública. Contribuindo para estudos de outro grupos de pesquisa, expandindo os conhecimentos sobre RF.

#### 5.4 PUBLICAÇÕES

PRODOSSIMO, F. C.; CHIDAMBARAM, C.; LOPES, H. S. Otimização da Detecção de Olhos em Imagens Faciais utilizando os Algoritmos Colônia de Abelhas Artificiais e Harmony Search. In: X CONGRESSO BRASILEIRO DE INTELIGÊNCIA COMPUTACIONAL, 2011. Fortaleza. Anais do X Congresso Brasileiro de Inteligência Computacional (CBIC 2011).

PRODOSSIMO, F. C.; CHIDAMBARAM, C.; LOPES, H. S. Proposta de uma Metodologia Híbrida para a Construção de um Banco de Imagens Faciais Normalizadas. In: VIII WORKSHOP DE VISÃO COMPUTACIONAL, 2012. Goiânia. Anais do VIII Workshop de Visão Computacional (WVC 2012).

PRODOSSIMO, F. C.; CHIDAMBARAM, C.; LOPES, H. S. Eye detection Approaches based on AdaBoost and SVM algorithms. In: IX WORKSHOP DE VISÃO COMPUTACIONAL, 2013. Niterói. (Artigo aceito para apresentação de pôster, a ser publicado).

## REFERÊNCIAS

- AKAMATSU, M. L. abd S.; KAMACHU, M.; GYOBA., J. Coding facial expressions with gabor wavelets. **Third IEEE International Conference on Automatic Face and Gesture Recognition**, p. 200 – 205, 1998.
- AMARAL, V. do et al. Normalização espacial de imagens frontais de face em ambientes controlados e não-controlados. São Caetano do Sul - SP, v. 1, n. 1, p. 87 – 99, 2009.
- AMARAL, V. do; THOMAZ, C. E. **Normalização Espacial de Imagens Frontais de Face**. São Caetano do Sul - SP, 2008. Relatório Técnico.
- CHAM, C. H. **The Extended M2VTS Database**. 2006. Disponível em: <<http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/>>.
- CHANG, K. I.; BOWYER, K. W.; FLYNN, P. J. Face recognition using 2d and 3d facial data. **ACM Workshop on Multimodal User Authentication**, p. 25–32, 2003.
- CHIDAMBARAM, C.; LOPES, H. S. A new approach for template matching in digital images using an artificial bee colony algorithm. **Proceedings of Nature & Biologically Inspired Computing World Congress**, p. 146–151, 2009.
- CHIDAMBARAM, C.; LOPES, H. S. An improved artificial bee colony algorithm for object recognition problem in complex digital images using template matching. **International Journal of Natural Computing Research**, v. 1, n. 2, p. 54–70, 2010.
- DORIGO, M.; COLORNI, A. The ant system: optimization by a colony of cooperating agents. **IEEE Transactions on Systems, Man and Cybernetics - Part B**, v. 26, p. 29–41, 1996.
- DUBUISSON, M. P.; JAIN, A. K. A modified hausdorff distance for object matching. **IEEE International Conference on Pattern Recognition**, p. 566 – 568, 1994.
- EBERHART, R.; KENNEDY, J. A new optimizer using particle swarm theory. **Sixth International Symposium on Micro Machine and Human Science**, p. 5, 1995.
- FACIAL IDENTIFICATION SCIENTIFIC WORKING GROUP. **Capture and Equipment Assessment for Facial Recognition Systems**. [S.l.], 2011. Versão 1.0.
- FACON, J. **Morfologia Matemática: Teoria e Exemplos**. Curitiba - PR: Editora Universitária Champagnat da Pontifícia Universidade Católica do Paraná, 1996. 320 p.
- FREIRE, D.; SANTANA, M. C.; DÉNIZ-SUÁREZ, O. Smile detection using local binary patterns and support vector machines. **Proceedings of the Fourth International Conference on Computer Vision Theory and Applications (VISAPP 2009)**, v. 1, p. 5 – 8, 2009.
- FREUND, Y.; SCHARPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. **Journal of Computer and System Sciences**, v. 55, p. 119 – 139, 1997.

- FRISCHHOLZ, R.; DIECKMANN, U. Bioid: a multimodal biometric identification system. **Computer**, v. 33, p. 64 – 68, 2000.
- GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing - Third Edition**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.
- GRIFFIN, P. **Face Recognition Format for Data Interchange**. Minnetonka - MN, 2003. Relatório Técnico.
- GROSS, R. Face databases. **Handbook of Face Recognition**, 2005.
- HANSEN, P.; MLADENOVIC, N. Variable neighborhood search: Principles and applications. **European Journal of Operational Research**, v. 130, p. 449–467, 2001.
- ISLAM, R. **Police Standard for Still Digital Image Capture and Data Interchange of Facial/Mugshot and Scar, Mark and Tattoo Images**. 2.0. ed. United Kingdom, Maio 2007. Manual.
- JACQUES, J. C. S. J. et al. A template-matching based method to perform iris detection in real-time using synthetic templates. **Proceedings of IEEE International Symposium on Multimedia**, p. 142–147, 2009.
- JESORSKY, O. et al. Robust face detection using the hausdorff distance. **Proceedings of International Conference on Audio- and Video-based Biometric Person Authentication**, p. 90 – 95, 2001.
- JIAN, W.; HONGLIAN, Z. Eye detection based on multi-angle template matching. **Proceedings of International Conference on Image Analysis and Signal Processing**, p. 241 – 244, 2009.
- KE, L.; HUANG, Y. Eye location based on dual-orientation gabor filters and templates. **Proceedings of International Congress on Image and Signal Processing**, p. 4, 2009.
- LEITE, B. de B. **Deteção de Olhos em Imagens com Faces Humanas**. 109 p. Dissertação (Mestrado) — Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2008.
- LORENA, A. C.; CARVALHO, A. C. P. L. F. de. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, XIV, p. 43 – 67, 2007.
- MARQUES FILHO, O.; VIEIRA NETO, H. **Processamento Digital de Imagens**. Rio de Janeiro: Brasport, 1999.
- MLADENOVIC, N.; HANSEN, P. Variable neighborhood search. **Computers and Operations Research**, v. 24, p. 1097 – 1100, 1997.
- NIST. **Best Practice Recommendation for the Capture of Mugshots - NIST Best Practices**. United States, 1997. Manual de Boas Práticas.
- NIST. **The FERET Database**. 2004. Disponível em: <<http://www.itl.nist.gov/iad/humanid/feret/>>.
- NIST. **Face Recognition Format for Data Interchange - Best Practices**. United States, 2007. Manual de Boas Práticas.

- NORDSTROM, M. M. et al. **The IMM face database - an annotated dataset of 240 face imagens**. [S.l.], 2004.
- OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on feature distributions. **Pattern Recognition**, v. 29, p. 51 – 59, 1996.
- OJALA, T.; PIETIKÄINEN, M.; MÄENPÄÄ, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 24, p. 971 – 987, 2002.
- OLIVEIRA, L. L. J. de; THOMAZ, C. E. **Captura e Alinhamento de Imagens: Um Banco de Faces Brasileiro**. [S.l.], 2005. Relatório Técnico.
- PAPAGEORGIOU, C. P.; OREN, M.; POGGIO, T. A general framework for object detection. **International Conference on Computer Vision**, p. 555 – 562, 1998.
- PERLIN, H. A. **Reconhecimento de Objetos em Imagens Digitais Utilizando Otimização por Enxame de Partículas**. Dissertação (Mestrado) — Universidade Tecnológica Federal do Paraná, Curso de Pós-Graduação em Engenharia Elétrica e Inforática Industrial, 2009.
- PERLIN, H. A.; LOPES, H. S.; CENTENO, T. M. Particle swarm optimization for object recognition in computer vision. **Proceedings of 21st International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems**, p. 11–21, 2008.
- PHILIPS, P. J.; RIZVI, S. A. The feret evaluation methodology for face-recognition algorithms. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, p. 1090 – 1104, 2000.
- POLLI, R. M. et al. A proposal for the hand palm identification, using local binary pattern. **International Journal of Advanced Engineering Sciences and Technologies**, v. 9, p. 302 – 309, 2011.
- PRODOSSIMO, F. C.; CHIDAMBARAM, C.; LOPES, H. S. Otimização da detecção de olhos em imagens faciais utilizando os algoritmos colônia de abelhas artificiais e harmony search. **Anais do X Congresso Brasileiro de Inteligência Computacional**, 2011.
- PRODOSSIMO, F. C.; CHIDAMBARAM, C.; LOPES, H. S. Proposta de uma metodologia para a construção de um banco de imagens faciais normalizadas. **Anais do VIII Workshop de Visão Computacional**, 2012.
- REISFELD, D.; YESHURUN, Y. Preprocessing of face images: Detection of features and pose normalization. **Computer Vision and Image Understanding**, v. 71, p. 413–430, 1998.
- RIZVI, S. A.; PHILIPS, P. J.; MOON, H. **The FERET Verification Testing Protocol for Face Recognition Algorithms**. [S.l.], 1998.
- RUBINFELD, M.; WILSON, C. Gray calibration of digital cameras to meet nist mugshot best practice. 1999.
- SCALABRIN, M. H.; PARPINELLI, R. S.; LOPES, H. S. Paralelização do algoritmo harmony search utilizando unidade de processamento gráfico. **Mecânica Computacional**, v. 29, p. 7109–7121, 2010.

SCHWENK, H.; BENGIO, Y. Boosting neural networks. **Neural Computation**, v. 12, p. 1869 – 1887, 2000.

VAPNIK, V. N. **The nature of Statistical learning theory**. New York: Springer-Verlag, 1995.

VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. **Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, v. 1, p. 511 – 518, 2001.

WANG, L.; HE, D. C. Texture classification using texture spectrum. **Pattern Recognition**, v. 23, p. 905 – 910, 1990.

WEBER, M. **Caltech Frontal Face Database**. 1999. Disponível em: <<http://www.vision.caltech.edu/html-files/archive.html>>.

WISKOTT, L. et al. Face recognition by elastic bunch graph matching. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 19, p. 775 – 779, 1997.

ZHAO, W. et al. Face recognition: A literature survey. **ACM Computing Surveys**, v. 35, p. 399–458, 2003.

## ANEXO A – TERMO DE ASSINATURAS PARA CAPTURA DAS IMAGENS FACIAIS PARA CONSTRUÇÃO DA BASE

### Autorização para Publicação de Imagens

Estamos Iniciando uma pesquisa integrada sobre reconhecimento facial (Projeto FR – Face Recognition) no CPGEI (Curso de Pós-Graduação em Engenharia Elétrica e Informática Industrial).

O objetivo deste estudo é a aquisição de imagens faciais para a construção de uma base de imagens.

A participação é voluntária. A identidade dos participantes será mantida através de códigos numéricos. Portanto, o nome e os demais dados não serão divulgados junto com as imagens.

Através deste, autorizo a publicação da minha imagem nos artigos, congressos e eventos científicos e outros eventos relacionados.

Nº	Nome	RG	E-mail	Idade	Sexo	Assinatura
181						
182						
183						
184						
185						
186						
187						
188						
189						
190						

**Figura 49: Termo de Autorização.**

**Fonte: Autoria Própria.**



## **ANEXO B – PUBLICAÇÕES**

X Congresso Brasileiro de Inteligência Computacional (CBIC'2011), 8 a 11 de Novembro de 2011, Fortaleza, Ceará  
© Sociedade Brasileira de Inteligência Computacional (SBIC)

## OTIMIZAÇÃO DA DETECÇÃO DE OLHOS EM IMAGENS FACIAIS UTILIZANDO OS ALGORITMOS COLÔNIA DE ABELHAS ARTIFICIAIS E *HARMONY SEARCH*

**Flávio das Chagas Prodossimo**

Universidade Tecnológica Federal do Paraná  
frovas@gmail.com

**Chidambaram Chidambaram**

Universidade do Estado de Santa Catarina  
chidambr@gmail.com

**Heitor Silvério Lopes**

Universidade Tecnológica Federal do Paraná  
heitorslopes@gmail.com

**Resumo** – A tarefa do reconhecimento facial geralmente inicia por detecção de olhos, que é considerada como a fase mais importante. De modo geral, a detecção de olhos é necessária para detecção e normalização de faces e ainda, para extração de características. O encontro dos olhos na imagem facial pode ser feita através de uso de *templates*. Porém, a imagem facial poderá ter várias partes similares. Portanto, o uso da técnica *Template Matching* (TM) e de algoritmos de otimização pode tornar mais eficiente a localização de olhos. Neste trabalho a busca é otimizada usando o algoritmo de *Harmony* (HS) e o algoritmo Artificial de Colônia de Abelhas (ABC). Considerou-se nos experimentos os fatores como escala e rotação da imagem, além da localização espacial (eixos x e y). Os resultados são comparados para avaliar o desempenho dos algoritmos discutidos neste trabalho. Com o trabalho foi possível notar que tanto o *Harmony Search* quanto o algoritmo de Colônia de Abelhas detectaram os olhos nas faces, algumas vezes com os olhos centralizados na imagem recorte.

**Palavras-chave** – Detecção de olhos, Correlação, *Harmony Search*, Colônia Artificial de Abelhas, Otimização.

**Abstract** – The task of face recognition begins with eye detection that is considered the most important phase. In general, the eye detection is needed for face detection and normalization and feature extraction. The matching of the eyes can be done through templates. However, the use of the technique *Template Matching* and optimization algorithms can make more efficient the eye location. In this work the search is optimized using the *Harmony Search* (HS) algorithm and the Artificial Bee Colony algorithm. It was considered in the experiments scale and rotation factors, in addition to spatial location (x and y axes). The results are compared to evaluate the performance of the algorithms discussed in this paper. With the work we became aware that the *Harmony Search* Algorithm and the Artificial Bee Colony detected eyes in the faces, sometimes with the eyes centered in the cutout image.

**Keywords** – Eye detection, Matching, *Harmony Search*, Artificial Bee Colony, Optimization.

### 1. INTRODUÇÃO

Com a evolução da Tecnologia da Informação, tem se tornando comum o reconhecimento de pessoas através de imagens obtidas através de câmeras de monitoramento ou imagens estáticas. Por questões de segurança, tem aumentado a demanda por sistemas automáticos para reconhecimento de faces, tanto na área pública quanto privada. Há muitos estudos e pesquisas que vêm sendo desenvolvidos pela comunidade científica com o objetivo de identificar pessoas usando somente a imagem da face. Para tanto, emprega-se técnicas e métodos científicos de diversas áreas, tais como Visão Computacional, Reconhecimento de Padrões e Inteligência Computacional.

De modo geral, o reconhecimento de pessoas através de faces ainda constitui uma tarefa desafiadora, pois as imagens que contém faces possuem alta variabilidade em tamanho, forma, cor e textura. Ainda mais, há necessidade de identificação de faces invariantes à posição, orientação ou variações de iluminação. Nas últimas décadas, inúmeras técnicas têm sido desenvolvidas para identificar faces numa base de imagens, por exemplo, usando as imagens de olhos como *templates* [1]. Desta forma, tem aumentado a demanda por métodos eficientes e robustos para a localização de faces humanas em imagens digitais, exigindo grande esforço computacional, técnicas inovadoras e algoritmos eficientes.

O reconhecimento facial requer diversas etapas [2]. Estas etapas incluem a detecção e normalização da face, extração de características e comparação com as imagens armazenadas em uma base. Dentre tais etapas, a mais importante é a detecção da

face propriamente dita. Esta fase é etapa essencial para as etapas seguintes. Neste contexto, o elemento facial mais importante para o reconhecimento são os olhos. A detecção dos olhos facilita o processo de normalização, que é necessária para padronizar as imagens.

Um processo bastante utilizado para encontrar padrões em imagens é o *Template Matching* (TM). O TM consiste em comparar a imagem de entrada e as imagens *templates* (padrões) da base através da correlação [3]. Neste trabalho a comparação é feita entre os *templates* de olhos predefinidos e a imagem de um olho extraída da imagem da face. Portanto, esta técnica leva à comparação exaustiva entre duas imagens [4], o que é computacionalmente custoso [3]. Para agilizar o processo de detecção de olhos através do TM e, ao mesmo tempo, reduzir o número de comparações, pode-se utilizar os algoritmos de otimização, neste caso, os algoritmos de Computação Evolucionária [5]. Dentre estes algoritmos podem ser citados: os Algoritmos Genéticos (AG), algoritmos baseados em Enxames de Partículas (*PSO-Particle Swarm Optimization*) e Colônia de Abelhas (*ABC-Artificial Bee Colony*), algoritmos baseados na música (*HS-Harmony Search*) [6], entre outros.

Neste trabalho, o objetivo é detectar os olhos em imagens faciais utilizando o TM através com algoritmos de otimização ABC e HS, analisando as vantagens e desvantagens de cada método.

As seções seguintes do artigo estão organizadas da seguinte forma: uma breve revisão sobre trabalhos relacionados e os principais técnicas e algoritmos utilizados no trabalho se encontram na Seção 2; o processo de detecção dos olhos é apresentado na Seção 3; os resultados são apresentados na Seção 4; e finalmente, a conclusão baseada nos experimentos e seus resultados é apresentada na Seção 5.

## 2. REVISÃO DA LITERATURA

Nesta seção são apresentados conceitos que foram utilizados neste projeto de pesquisa e alguns trabalhos correlatos desenvolvidos por outros pesquisadores e que serviram de embasamento teórico para o desenvolvimento deste trabalho.

### 2.1 Reconhecimento Facial

O reconhecimento facial (RF) vem sendo cada vez mais utilizado nas mais diversas áreas e se tornou um importante artefato para a segurança. O reconhecimento facial na área policial é utilizado para identificar as pessoas que praticam algum tipo de ato ilícito, em especial aquelas que já têm ficha policial cuja face se encontra na base de imagens. Além disto, o reconhecimento facial poderia ser utilizado na área de segurança privada como artefato de acesso a lugares restritos a algumas pessoas.

O RF abrange diversas etapas, desde a captura das imagens para treinamento do sistema até a fase de reconhecimento em si. Dentre estas etapas estão: detecção da face (ou elementos específicos da face, como os olhos, assim como é abordado neste trabalho), normalização da face, extração de características e o próprio reconhecimento [7, 8].

A detecção dos olhos é a mais importante de todas as etapas, pois a maioria das etapas seguintes depende da localização do olho. Portanto, é necessário localizar os olhos primeiro, assim como seu grau de rotação e sua variação em relação à escala para que a face seja normalizada. A extração de características, geralmente, é feita em regiões de interesse como olhos, nariz e boca [8].

### 2.2 Template Matching

Uma das técnicas que pode ser utilizada para realizar a detecção de olhos, ou outro objeto de interesse na imagem, é o *Template Matching* (TM). O principal objetivo do TM é encontrar um padrão em algum lugar de uma imagem, fazendo a correlação entre as duas imagens. O padrão que se deseja encontrar é conhecido como Imagem de Referência (IR) e a imagem na qual o padrão deve ser localizado é chamada de Imagem de Paisagem (IP) [9]. Em particular, neste trabalho a IR é o olho, e aqui chamada de Imagem-Olho de referência (IOR), enquanto que a IP da face humana é chamada de Imagem-Face (IF). Esta terminologia será utilizada ao longo deste artigo.

O processo de TM é feito usando o recorte de imagem feita na IF (no mesmo tamanho da IOR) e a IOR. Na IF, cada pixel pode ser uma coordenada central  $(x, y)$  do objeto a ser reconhecido. A vantagem de utilizar uma metodologia como esta é que não é necessário a segmentação da imagem. Uma desvantagem do TM é que o padrão que se deseja encontrar pode estar em qualquer lugar da IF, sendo, portanto, necessário buscar em toda a IF. Além disto, o padrão pode ter sofrido rotação e/ou escala, o que torna processo de busca complexo e custoso computacionalmente, pois além das coordenadas  $x$  e  $y$  da localização planar, deve-se determinar também os fatores de rotação e escala. De modo geral, métodos baseados em correlação como TM são muito dependentes de iluminação, rotação e escala [10].

Considerando todos estes fatores, o número de recortes possíveis numa imagem (espaço de busca) pode ser intratavelmente grande. Para realizar uma busca mais eficiente em tempo aceitável, utilizou-se os algoritmos de otimização ABC e HS implementados por [11] e [12], respectivamente.

### 2.3 Algoritmos de Otimização

Os algoritmos de otimização são um conjunto de técnicas de busca por possíveis soluções para problemas complexos. Muitas vezes tais algoritmos são inspirados na natureza. Para esta classe de algoritmos, utilizam-se populações de indivíduos no processo de busca, onde cada indivíduo pode ser uma possível solução para o problema. A ideia principal é evoluir, de alguma maneira, os indivíduos através de operações entre eles e da aplicação de princípios da evolução natural. A qualidade do indivíduo, enquanto

X Congresso Brasileiro de Inteligência Computacional (CBIC'2011), 8 a 11 de Novembro de 2011, Fortaleza, Ceará  
© Sociedade Brasileira de Inteligência Computacional (SBIC)

solução do problema, é calculada através de uma função chamada de *fitness*, que mede a adequação do indivíduo. Assim sendo, o algoritmo é executado algumas vezes procurando por uma solução que melhor represente a resposta do problema [5].

Diversas fontes de inspiração na natureza foram utilizadas para o desenvolvimento dos algoritmos evolucionários. Um dos mais conhecidos e utilizados é o Algoritmo Genético (AG) proposto por [13], baseado na codificação genética das espécies. Neste algoritmo uma possível solução é análoga a um cromossomo biológico. Nos AGs a população inicial é formada por vários indivíduos (cromossomos) que são submetidos a evolução, sendo que uma parte da população é selecionada para próxima geração, gerando filhos através de cruzamento e mutação, os principais operadores do AG. Karungaru [14] propôs um trabalho sobre reconhecimento facial baseado no *Template Matching* otimizado com Algoritmos Genéticos. Este autor utilizou um *template* em "T" formado por olhos, nariz e boca para realizar o RF.

O comportamento de grupos de animais também foi fonte de inspiração para Eberhart et al. [15], que propuseram algoritmos que levam o nome de algoritmos de inteligência coletiva. Os principais algoritmos baseados nesta ideia são os de colônia de formigas (ACO - *Ant Colony Optimization*), colônia de abelhas (ABC - *Artificial Bee Colony*) [16] e por enxame de partículas (PSO - *Particle Swarm Optimization*) [15]. O ACO usa a comunicação através de trilhas de feromônio para indicar caminhos para fontes de alimentos. O PSO utiliza a ideia de boas experiências que são comunicadas entre indivíduos, fazendo com eles busquem conhecimento. Perlin [9, 17] utilizou PSO para otimizar a técnica de TM para o reconhecimento de objetos. O ABC utiliza a busca e coleta por alimento como ideia central. Este algoritmo foi usado por [11] para otimizar o processo de reconhecimento de objetos presentes num imagem de paisagem usando o TM. Da mesma forma, o ABC também foi utilizado em estudos de detecção de frís por [18].

A música também foi utilizada para formular metodologias inteligentes de busca, como é o exemplo do *Harmony Search* (HS) proposto por [6]. O HS utiliza a ideia do improviso de novos sons musicais para o encontro de uma harmonia perfeita. Scalabrin [12] utilizou o HS para otimizar processos de *benchmark*.

Neste trabalho, ABC e HS foram utilizados para buscar com mais eficiência no espaço de busca nas imagens. As seções a seguir explicam o funcionamento destes algoritmos.

#### 2.4 Harmony Search

O primeiro algoritmo de otimização que foi utilizado neste trabalho é o Harmony Search (HS), conforme proposto por [6], tendo como inspiração o arranjo de harmonias de um trio de jazz. A harmonia musical é um conjunto de sons que seja agradável ao ouvido humano. As harmonias são combinações de várias ondas de sons que possuem diferentes frequências [6].

Assim como outros algoritmos de otimização que buscam a melhor solução determinada por uma função objetivo, o HS busca a melhor harmonia de acordo com uma estimativa estética. Esta estimativa estética é dada pelos sons emitidos através dos instrumentos dos músicos e os sons podem ser melhorados através de prática. No HS esta estimativa pode ser comparada com o *fitness* de um Algoritmo Genético. Uma harmonia é melhorada ao longo das iterações buscando alcançar um bom resultado, assim como acontece com o *fitness* dos AGs. Os passos do algoritmo *Harmony Search* são mostrados a seguir:

1. Inicializar o problema: Parâmetros, Função de Avaliação e Critério de Parada. Dentre os parâmetros estão: Número de ciclos, número de variáveis (músicos), tamanho da Memória Harmônica, taxa de consideração da Memória Harmônica (HMCR), ajuste fino (PAR) e intervalos de variáveis.
2. Inicializar a Memória Harmônica (HM - *Harmony Memory*): Tamanho e harmonias aleatórias para cada posição da memória dentro do intervalo pré-definido.
3. Improvisar uma nova harmonia. A nova harmonia é composta por harmonias presentes na HM. Os parâmetros HMCR e PAR são responsáveis pelo controle de Busca Global e Local.
4. Atualizar Memória Harmônica. A nova harmonia é avaliada utilizando o elitismo. Se a nova harmonia for melhor que a melhor harmonia da HM, ela é substituída na HM, caso contrário a harmonia melhor é mantida.
5. Critério de parada, geralmente, é determinado pelo número máximo de ciclos. Caso contrário, retorna ao Passo 3.

Os principais parâmetros deste algoritmo são: o número de execuções e o número de músicos. O primeiro é o número de gerações que o algoritmo vai executar para encontrar a harmonia perfeita e o segundo, representa o número de variáveis presentes no problema. O HS sofre grande influência destes parâmetros na etapa de improviso 3.

Outros parâmetros são o HMCR (*Harmony Memory Considering Rate*) e o PAR (*Pitch Adjusting Rate*). O HMCR é responsável por definir a busca global (quanto maior o HMCR menor a busca global) e o PAR é responsável por ajustes finos para realizar a busca local. O HMCR é utilizado quando não existe partes da solução global na Memória Harmônica. Quando este parâmetro é ativado o *Harmony Search* encontra notas de forma aleatória dentre as possíveis notas para o problema. O HMCR pode variar entre 0 e 1, por exemplo, quando o valor for igual a 0,8 significa que o HS escolhe um valor para certa variável da HM com 80% de probabilidade. Por outro lado, o PAR é uma taxa de ajuste do passo para que o algoritmo não fique parado em um ponto de máximo/mínimo local. Este valor é o deslocamento de valores vizinhos dentro de um intervalo. Quando o valor do PAR é 0,1 significa que a probabilidade de se escolher um valor vizinho é de 10%. [6, 12]

## 2.5 Artificial Bee Colony

Os principais componentes que formam o algoritmo ABC são as fontes de comida, as abelhas empregadas e as abelhas desempregadas, estas podem ser classificadas em escoteiras e observadoras [19]. As abelhas empregadas avaliam as fontes de alimentos levando em consideração várias propriedades relacionadas as fontes encontradas tais como a proximidade à colônia, qualidade e quantidade de néctar ou pólen e as facilidades para extrair o alimento, se é fácil ou difícil [20]. Estas informações são levadas para a colmeia e transmitidas às abelhas observadoras através da dança. As abelhas escoteiras, por sua vez, procuram sempre aleatoriamente por novas fontes de comida, enquanto que as observadoras, usando a informação obtida da dança, decidem sobre a fonte que será explorada.

O processo aleatório de busca por alimentos se inicia numa colônia, por abelhas empregadas que são enviadas em várias direções simultaneamente para cobrir uma grande área de busca. É importante destacar que as abelhas possuem capacidade para localizar as fontes longe da colônia e memória para retornar à colônia sem se perder no caminho. Assim, também, elas têm capacidade para ajustar o padrão de busca e coleta de alimentos de acordo com as mudanças ambientais. Durante a busca, elas selecionam as melhores fontes com abundância de néctar ou pólen [11].

O algoritmo ABC é mostrado simplificadaamente a seguir:

1. Inicializar a População de tamanho SN.
2. Repetir até MCN:
  - (a) Posicionar abelhas empregadas no espaço de busca.
  - (b) Posicionar abelhas observadoras nas fontes de acordo com quantidade de néctar (transmitida pela dança).
  - (c) Determinar fontes de alimento a serem abandonadas.
  - (d) Enviar abelhas escoteiras para procurar novas fontes.
  - (e) Memorizar a melhor fonte que foi encontrada até o momento.
3. Fim.

Baseado no levantamento feito sobre o comportamento de abelhas, definiu-se vários parâmetros para o Algoritmo. Em ABC, a posição de uma fonte de alimento pode representar uma possível solução (neste trabalho, as coordenadas centrais da imagem cortada) para um problema de otimização e a qualidade e a quantidade de néctar de uma fonte pode corresponder a qualidade de uma solução que, por sua vez, pode ser considerada como *fitness*. O processo é executado por um Número Máximo de Ciclos (MCN - *Maximum Cycle Number*), ou até encontrar o critério de parada. A população de abelhas empregadas e observadoras presentes na colônia é representado por SN.

Na etapa 1 a população é iniciada com SN abelhas e as fontes de alimentos aleatórias são escolhidas. Na etapa 2a as abelhas empregadas são posicionadas nestas fontes. As abelhas voltam à colmeia e transmitem, através da dança, informações para as abelhas observadoras. As abelhas empregadas voltam para as fontes de alimento para selecionar novas fontes e novamente transmitir informações para as observadoras. Na etapa 2b as observadoras decidem, através de informações obtidas, para qual fonte de alimento elas irão, então elas são posicionadas nestas fontes e também procuram por novas fontes. Quando uma fonte se torna baixa ela é abandonada, etapa 2c. Em cada ciclo pelo menos uma abelha escoteira vai à procura de novas fontes de alimento para substituir a que foi abandonada (2d). A melhor fonte encontrada é memorizada (2e) e o processo começa novamente da etapa 2a até o critério de parada ser atingido ou até o valor de MCN.

## 3 PROCESSO DE DETECÇÃO DE OLHOS

O processo de detecção de olhos foi feito com a aplicação do método de correlação, TM. Considerando-se a localização de olhos como um problema de otimização, os algoritmos HS e ABC podem ser utilizados para encontrar o ponto de maior similaridade entre a IOR e a imagem cortada da IF baseado nos quatro parâmetros ( $x$ ,  $y$ ,  $s$ ,  $(\theta)$ ). A solução para esse problema é composta por 4 valores, sendo dois deles a localização espacial (eixos  $x$  e  $y$ ) da imagem de recorte. Esses valores são variáveis de 0 até o tamanho da imagem (largura e altura). Os demais valores são os fatores de rotação ( $\theta$ ) e escala ( $e$ ). O valor de rotação pode variar entre  $-\pi/2$  e  $\pi/2$  e a escala entre 0,5 e 1,5.

Neste trabalho, para evitar uma busca exaustiva usando o TM, os algoritmos de otimização HS e ABC foram utilizados para encontrar os olhos na IF usando um *template* de olho. No HS as variáveis  $x$ ,  $y$ ,  $\theta$  e  $e$  são representadas por uma harmonia e no ABC, por uma abelha. As informações contidas tanto na harmonia quanto na abelha servem para recortar a imagem na IF que posteriormente é utilizada no TM.

Para testes iniciais foram utilizadas imagens faciais controladas da base *FEI Face Database* [21] com o tamanho de 640 x 480 pixels. Em seguida, foram utilizadas imagens capturadas para realização deste estudo, com o tamanho de 1400 X 1050 pixels. Algumas destas imagens podem ser vistas na Figura 1. As IORs foram extraídas manualmente das imagens capturadas para a realização dos testes, cada uma possuindo 143 X 68 pixels. Como as IORs foram extraídas manualmente já foram estabelecidos os valores esperados da posição ( $x,y$ ) do centro do olho dos recortes, onde a imagem do olho ficasse centralizada.

O processo de detecção dos olhos parte do princípio de que uma IOR sai em busca de uma parte de imagem, na IF, que faça o melhor *matching* com a IOR. Este processo de busca de melhor correlação é auxiliado, neste trabalho, pelo algoritmo HS ou pelo ABC. O resultado mostra as coordenadas ( $x$ ,  $y$ ) da posição central do olho, além da rotação ( $\theta$ ) e escala ( $e$ ) da IF. Sabe-se que o ótimo global seria o *matching* do olho (IOR) da própria pessoa da IF.

X Congresso Brasileiro de Inteligência Computacional (CBIC'2011), 8 a 11 de Novembro de 2011, Fortaleza, Ceará  
 © Sociedade Brasileira de Inteligência Computacional (SBIC)

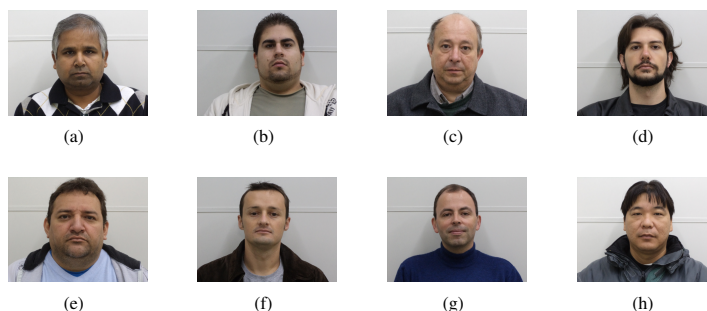


Figura 1: Imagens de Referência

#### 4. EXPERIMENTOS E RESULTADOS

Para realizar os testes, os algoritmos foram implementados em linguagem C, utilizando as funções do OpenCV. Para os testes do algoritmo HS os parâmetros foram iniciados com base nos estudos de Geem [22] e os parâmetros do ABC foram iniciados de acordo com sugestões de Karaboga e Akay [20]. Exemplos de testes estão descritos nesta seção e os resultados são apresentados em tabelas. A IFs utilizadas nos testes são apresentadas na Figura 1.

Os parâmetros para o algoritmo ABC para os primeiros testes (Teste1) foram: MCN igual a 200 e SN igual a 10, onde uma abelha é a escoteira. Em seguida foram feitos testes com MCN igual a 20 e SN igual a 100 (Teste2). Para cada teste o algoritmo foi executado 30 vezes. Na Tabela 1  $\bar{dE}$  é a média da distância Euclidiana dos pontos encontrados em relação aos pontos reais (manualmente determinados no centro dos olhos). Os pontos encontrados também são as coordenadas (x, y) do centro da imagem. Os campos descritos como Avaliações representam o número médio de vezes que o *fitness* foi calculado para cada rodada dos algoritmos Para estes testes as IFs estão sem rotação e mudança de escala.

Tabela 1: Soluções encontradas pela execução do TM com ABC (valores médios)

IR	$\bar{dE}$ (Teste1)	Avaliações	$\bar{dE}$ (Teste2)	Avaliações
1(a)	40 ± 29	9900 ± 961	9 ± 3	5324 ± 907
1(b)	158 ± 40	10160 ± 2285	51 ± 50	5957 ± 1128
1(c)	64 ± 50	10400 ± 2581	11 ± 23	5421 ± 1157
1(d)	76 ± 69	10305 ± 2219	6 ± 4	5863 ± 1704
1(e)	131 ± 118	10350 ± 2094	7 ± 11	5395 ± 1547
1(f)	66 ± 66	10735 ± 2575	4 ± 3	5354 ± 1141
1(g)	78 ± 81	10670 ± 2209	11 ± 27	5729 ± 1175
1(h)	45 ± 46	10620 ± 1930	15 ± 4	5538 ± 1396
Média		10392		5572

Como é possível observar nesta tabela (1) os resultados do Teste2 foram melhores. Mais valores do Teste2 do que do Teste1 se aproximaram do esperado. O Teste1 mostra que um número menor de abelhas faz com que o algoritmo não seja tão eficiente, pois não é feita uma busca global tão ampla pelo espaço de busca antes de ser feita uma busca mais localizada. Observando os dois testes é possível verificar que a segunda IF (1(b)) é a que mostrou maior distância Euclidiana. Isso ocorreu porque o algoritmo identificou o cabelo como um ponto de mínimo local na maioria das rodadas, por ele conter muitos pixels semelhantes com a IOR, não conseguindo chegar ao ótimo global.

Alguns testes foram realizados para o HS com os parâmetros definidos como segue. No Teste1 o HMCR foi igual a 0,1 e o PAR igual a 0,9, caracterizando uma busca aleatória. No Teste2 o HMCR foi definido como 0,9 e o PAR 0,2. Os resultados destes testes, também sem rotação e escala, são, mostrados na Tabela 2.

A Tabela 2 mostra que em ambos os testes o algoritmo HS foi eficiente. Porém, no primeiro foi necessário um número maior de avaliações para encontrar a IOR, devido às configurações de parâmetros, levando o algoritmo a privilegiar uma busca mais global. No segundo teste o olho não foi detectado na imagem 1(e) e, sim, o queixo do indivíduo, pois ele está com um pouco de sombra em um dos olhos, fazendo com que o algoritmo encontrasse parte da barba, também sombreada, pela similaridade dos pixels.

Testes com as IFs sofrendo rotação e escala também foram feitos. Todas as imagens foram submetidas a uma redução de 15%, ou seja, a escala esperada pela execução do algoritmo é 0,85. As 4 primeiras imagens foram submetidas a uma rotação de 20° e as outras de -20°. Sendo assim o valor esperado é 0,35 e -0,35, 20° e -20° em radianos respectivamente. Para estes testes o algoritmo foi executado 30 vezes. A seguir são mostrados os resultados das execuções do TM com ABC (Tabela 3) e com o HS

X Congresso Brasileiro de Inteligência Computacional (CBIC'2011), 8 a 11 de Novembro de 2011, Fortaleza, Ceará  
 © Sociedade Brasileira de Inteligência Computacional (SBIC)

Tabela 2: Soluções encontradas pela execução do TM com HS (valores médios)

<i>IR</i>	$\bar{dE}$ (Teste1)	Avaliações	$\bar{dE}$ (Teste2)	Avaliações
1(a)	7 ± 1	57432 ± 25568	7 ± 0, 2	47905 ± 24317
1(b)	6 ± 1	66190 ± 25931	5 ± 0, 2	23016 ± 24361
1(c)	6 ± 1	67460 ± 21914	7 ± 0, 3	46585 ± 33214
1(d)	5 ± 1	69560 ± 18214	5 ± 0, 5	25023 ± 16400
1(e)	3 ± 1	60150 ± 27335	251 ± 220	9859 ± 14523
1(f)	3 ± 1	67514 ± 27618	3 ± 0, 2	20701 ± 17739
1(g)	3 ± 1	73519 ± 17568	4 ± 0, 2	18544 ± 15584
1(h)	15 ± 1	60481 ± 23520	15 ± 0, 2	10513 ± 12762
Média		65288		25268

(Tabela 4). Nas tabelas são mostradas a média e desvio padrão das distâncias euclidianas dos valores obtidos com os esperados, ainda é possível ver a rotação e escala obtidos com cada IF.

Tabela 3: Soluções encontradas pela execução do TM com ABC para as IFs com rotação e escala (valores médios)

<i>IR</i>	$\bar{dE}$	$\theta$	$e$	Avaliações
1(a)	8 ± 0, 5	0, 36 ± 0, 01	0, 85 ± 0, 01	48476 ± 2533
1(b)	129 ± 9	-0, 97 ± 0, 07	0, 51 ± 0, 03	51598 ± 8951
1(c)	18 ± 36	0, 38 ± 0, 17	0, 83 ± 0, 08	48164 ± 9574
1(d)	2 ± 1	0, 33 ± 0, 04	0, 85 ± 0, 02	50224 ± 8652
1(e)	14 ± 52	-0, 35 ± 0, 02	0, 83 ± 0, 05	49300 ± 3698
1(f)	10 ± 22	-0, 35 ± 0, 02	0, 85 ± 0, 08	42092 ± 10917
1(g)	5 ± 1	-0, 36 ± 0, 01	0, 85 ± 0, 01	50636 ± 7295
1(h)	11 ± 2	-0, 35 ± 0, 01	0, 86 ± 0, 03	48714 ± 8821
Média				48650

Os parâmetros para este teste foram definidos como 100 e 200 para número de abelhas e número de gerações, respectivamente. Nesta tabela (3) pode-se notar que os olhos foram encontrados nas IFs, porém não totalmente centralizados. Com exceção da IF 1(b), nesta imagem o algoritmo está encontrando o cabelo ao lado esquerdo do olho. Nas imagens 1(c), 1(e) e 1(f) o desvio padrão foi grande pelo fato de em uma das 30 rodadas a distância Euclidiana ter sido maior que 100, caracterizando que não foi encontrado o olho. Porém, nas outras 29 rodadas o algoritmo foi eficiente.

Tabela 4: Soluções encontradas pela execução do TM com HS para as IFs com rotação e escala (valores médios)

<i>IR</i>	$\bar{dE}$	$\theta$	$e$	Avaliações
1(a)	163 ± 185	0, 12 ± 0, 3	0, 88 ± 0, 02	86930 ± 18365
1(b)	139 ± 165	0, 61 ± 0, 7	0, 90 ± 0, 25	71290 ± 31340
1(c)	20 ± 36	0, 18 ± 0, 5	0, 81 ± 0, 1	83650 ± 18493
1(d)	59 ± 96	0, 37 ± 0, 1	0, 77 ± 0, 1	88367 ± 13756
1(e)	115 ± 145	-0, 08 ± 0, 5	0, 88 ± 0, 22	92431 ± 7646
1(f)	129 ± 117	0, 31 ± 0, 6	0, 73 ± 0, 15	75737 ± 27493
1(g)	5 ± 1	-0, 37 ± 0, 02	0, 83 ± 0, 04	90917 ± 8676
1(h)	11 ± 1	-0, 34 ± 0, 04	0, 84 ± 0, 03	90100 ± 9918
Média				84927

Os resultados equivalentes à tabela 4 foram obtidos com HMCR igual a 0,7 e o PAR igual a 0,5. Com a execução do HS com esta configuração é possível notar que o algoritmo não foi eficiente. Apesar disso, as ultimas 2 imagens tiveram o olho detectado e em algumas imagens o olho foi detectado em menos da metade das execuções. Porém, as médias não ficaram de acordo com o esperado.

## 5. CONCLUSÃO

A proposta deste trabalho foi desenvolver um método de busca dos olhos numa imagem de face humana. A técnica utilizada foi o *Template Matching*, baseada na correlação entre imagens. Para otimizar esta técnica foram utilizados algoritmos de Computação Evolucionária, um baseado na música, o *Harmony Search* e outro baseado no comportamento de abelhas, o algoritmo de Colônia Artificial de Abelhas.

Com a execução dos testes foi possível notar que ambos os algoritmos foram eficientes para a detecção dos olhos. Em alguns casos o ABC foi melhor do que o HS, como é o exemplo das imagens sofrendo rotação e translação. O HS chegou a resultados mais próximos do esperado para as imagens sem rotação e escala. Porém, o número médio de avaliações foi entre 2 e 5 vezes maior que o ABC.

Com este trabalho foi possível notar que muitas vezes pequenos detalhes podem dificultar a detecção dos olhos. Como ocorreu nos testes, quando existia sombra nos olhos o algoritmo encontrava similaridade dos pixels com cabelo ou barba. Outro ponto que dificultou a detecção foi em função dos vários ótimos locais que o algoritmo interpreta, como exemplo cabelos, sobrancelhas e barba. Para encontrar o olho é necessário explorar globalmente o espaço de busca procurando ótimos locais, pois não existe uma tendência para o ótimo global (olho).

De modo geral, o algoritmo ABC foi mais eficiente que o HS. Apesar do HS ter detectado os olhos na maioria das execuções, em algumas não detectava, e ainda, o número médio de avaliações foi maior que do algoritmo ABC. O ABC realizou uma maior busca global no início da execução para depois uma busca localizada, fazendo com que os olhos fossem detectados em situações adversas, com rotação e escala.

E ainda, a partir da detecção de um dos olhos, é possível estimar a localização de outros componentes faciais como o nariz e a boca. Assim, é possível detectar a face por completo. Além disso, utilizando-se de este tipo de algoritmos de otimização, é possível realizar, por exemplo, a normalização de faces ou até o reconhecimento facial.

## 6. AGRADECIMENTOS

Aos voluntários que participaram da captura das imagens faciais. E ao colega Marlon Henrique Scalabrin por disponibilizar o algoritmo *Harmony Search* para o desenvolvimento desta pesquisa.

## REFERÊNCIAS

- [1] L. Ke and Y. Huang. "Eyes Location based on Dual-orientation Gabor Filters and Templates". *2nd International Congress on Image and Signal Processing*, pp. 1–4, 2009.
- [2] K. I. Chang, K. W. Bowyer and P. J. Flynn. "Face Recognition Using 2D and 3D Facial Data". *ACM Workshop on Multimodal User Authentication*, pp. 25–32, 2003.
- [3] L. Cole, D. Austin and L. Cole. "Visual Object Recognition using Template Matching". *In Proceedings of Australian Conference on Robotics and Automation*, 2004.
- [4] R. Brunelli. *Template Matching Techniques in Computer Vision: Theory and Practice*. Wiley, 2009.
- [5] T. Bäck, U. Hammel and H.-P. Schwefel. "Evolutionary Computation: Comments on the History and Current State". *IEEE Transactions on Evolutionary Computation*, vol. 1, pp. 3–17, 1997.
- [6] Z. W. Geem, J. Kim and G. V. Hoon, Longanathan. "A New Heuristic Optimization Algorithm: Harmony Search". *Simulation. Transactions of the Society for Modeling and Simulation International*, vol. 76, no. 2, pp. 60–68, 2001.
- [7] D. Reissfeld and Y. Yeshurun. "Preprocessing of Face Images: Detection of Features and Pose Normalization". *Computer Vision and Image Understanding*, vol. 71, pp. 413–430, 1998.
- [8] W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld. "Face Recognition: A Literature Survey". *ACM Computing Surveys*, vol. 35, pp. 399–458, 2003.
- [9] H. A. Perlin, H. S. Lopes and T. M. Centeno. "Particle Swarm Optimization for Object Recognition in Computer Vision". *21st international conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 11–21, 2008.
- [10] L. O. Marin. "Métodos Estatísticos no Reconhecimento de Faces". *Revista Eletrônica de Sistemas de Informação (RESI)*, vol. 8, no. 2, 2006.
- [11] C. Chidambaram and H. S. Lopes. "A New Approach for Template Matching in Digital Images Using an Artificial Bee Colony Algorithm". *Nature & Biologically Inspired Computing World Congress*, pp. 146–151, 2009.
- [12] M. H. Scalabrin, R. S. Parpinelli and H. S. Lopes. "Paralelização do Algoritmo Harmony search utilizando Unidade de Processamento Gráfico". *Mecânica Computacional*, vol. 29, pp. 7109–7121, 2010.



X Congresso Brasileiro de Inteligência Computacional (CBIC'2011), 8 a 11 de Novembro de 2011, Fortaleza, Ceará  
© Sociedade Brasileira de Inteligência Computacional (SBIC)

- [13] J. H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan Press, 1975.
- [14] S. Kaungaru, M. Fukumi and N. Akamatsu. “Face Recognition using Genetic Algorithm based Template Matching”. *International Symposium on Communications and Information Technologicis*, vol. 2, pp. 1252–1257, 2004.
- [15] R. C. Eberhart, Y. Shi and J. Kenedy. *Swarm Intelligence*. Morgan Kaufmann, San Francisco, 2001.
- [16] D. Karaboga. “An idea based on Honey Bee Swarm for Numerical Optimization”. Technical Report TR-06, Erciyes University - Engineering Faculty, October 2005.
- [17] H. A. Perlin. “Reconhecimento de Objetos em Imagens Digitais Utilizando Otimização por Enxame de Partículas”. Masters thesis, Universidade Tecnológica Federal do Paraná, 2009.
- [18] J. C. S. J. Junior, J. L. Moreira, A. Braun and S. R. Musse. “A Template-matching Based Method to Perform Iris Detection in Real-time Using Synthetic Templates”. *11th IEEE International Symposium on Multimedia*, pp. 142 – 147, 2009.
- [19] V. Tereshko and A. Loengarov. “Collective Decision-Making in Honey Bee Foraging Dynamics”. *Coomputing and Information Systems*, vol. 9, no. 3, pp. 1–7, 2005.
- [20] D. Karaboga and B. Akay. “A comparative study of Artificial Bee Colony algorithm”. *Applied Mathematics and Computation*, vol. 214, pp. 108–132, 2009.
- [21] L. L. de Oliveira Junior and C. E. Thomaz. “Captura e Alinhamento de Imagens: Um Banco de Faces Brasileiro”. Final report, Centro Universitário da FEI, 2005.
- [22] Z. W. Geem. *Recent Advances in Harmony Search Algorithm*. Springer, Berlin, 2010.

## Proposta de uma Metodologia para a Construção de um Banco de Imagens Faciais Normalizadas

Flávio das Chagas Prodossimo  
Universidade Tecnológica Federal do Paraná  
Av. Sete de Setembro, 3165  
frovas@gmail.com

Chidambaram Chidambaram  
Universidade do Estado de Santa Catarina  
Rua Luiz Fernando Hastreiter, 180  
chidambr@gmail.com

Heitor Silvério Lopes  
Universidade Tecnológica Federal do Paraná  
heitorslopes@gmail.com

### Abstract

*The area of computer vision is one of the most studied and is the basis for studies of face recognition. This area requires facial images under different conditions. Thus, the objective of this paper is to show the methodology applied to capture facial images which will be used in the experiments of various applications, for example, standardization, detection and face recognition. In this work, equipment configuration, the image measures and different poses for the capture of face images are presented. Several face images were captured and at then they were normalized according to international standards. From this work, it can be noted that predefined procedure and standards are important to capture images so that face recognition and other related tasks can be effectively performed.*

### 1. Introdução

Com surgimento de novas tecnologias e o aumento da preocupação nas questões de segurança, a identificação e o reconhecimento de pessoas usando imagens passou a ser uma das tarefas mais comuns e, ao mesmo tempo, se tornou essencial tanto na área de segurança privada quanto na pública. Para construção de sistemas de reconhecimento são necessárias várias imagens em diversas condições. Portanto, a construção de base de imagens se torna necessária a fim de realizar os experimentos utilizando diferentes métodos da área de reconhecimento facial.

No presente trabalho, o objetivo é reunir as informações necessárias para captura de imagens faciais com base em documentos nacionais e internacionais e, a partir disto, propor uma metodologia para capturar e normalizar estas imagens.

O estabelecimento de padrões tem como objetivo facilitar a análise detalhada e a verificação de imagens. Isso

pode auxiliar tanto os examinadores da segurança pública quanto a identificação através de métodos computacionais. Esses padrões objetivam definir um conjunto de orientações e recomendações que farão com que as imagens capturadas tenham uma qualidade mínima e suficiente para a verificação e a identificação de imagens faciais.

O trabalho também apresenta a metodologia que foi utilizada na construção de uma base de imagens faciais. No decorrer do documento é possível ver a forma como o ambiente deve ser organizado com os equipamentos para que as imagens sejam capturadas. Está descrito, também, como as imagens são capturadas, as rotações e translações da face, variações de iluminação, e a sequência em que isto é feito. Essas imagens estão sendo utilizadas em estudos de processamento facial, tais como normalização, detecção e reconhecimento.

Outra tarefa importante tratada neste trabalho é a normalização das imagens. A normalização das imagens inclui ajuste da rotação, escala, posicionamento da face, iluminação e cor do fundo da imagem [9]. Esse processo consta da padronização das imagens seguindo as padrões de normalização.

A principal motivação para realizar este estudo é a de ter encontrado somente um estudo semelhante em âmbito nacional [3]. Outra motivação que levou a este estudo é a falta de bases de imagens faciais que possuam imagens coloridas e com resolução grande, acima das resoluções das bases existentes e acima do indicado em [7] que é de 768x1024 pixels. Segundo o documento da NPIA [7], para que o reconhecimento automático seja implementado amplamente, o mais provável é que os padrões atuais devem ser atualizados, iniciando pelo aumento de resolução, captura de múltiplas imagens em vários ângulos e até as imagens 3D. Foram pesquisadas diversas bases na internet e no estudo de Gross [6], mas não foi encontrada nenhuma base pronta

com imagens nas condições esperadas. Portanto, decidiu-se a construção de uma base com condições específicas alinhadas às necessidades do projeto de reconhecimento facial que está em desenvolvimento.

## 2. Fundamentação Teórica

Para estudo de técnicas de visão computacional é necessário capturar imagens em diversas condições distintas. Para projetos de reconhecimento facial as bases de imagens devem ser compostas de várias posições das faces de pessoas, com isso é possível testar algoritmos de reconhecimento. A partir do momento em que as imagens estão disponíveis, existem etapas importantes para o reconhecimento facial. Uma delas é a normalização de imagens, que pode ser entendida como um pré-processamento das faces para utilizar os métodos de reconhecimento. Tanto para a captura quanto para a normalização de imagens faciais existem normas a ser seguidas [7] e [9].

Várias bases de imagens faciais foram analisadas para se obter informações sobre as imagens disponíveis para realização de reconhecimento facial, as bases mais relevantes podem ser vistas na Tabela 1. Levando em consideração estes dados e, através do estudo de Gross [6], no decorrer do artigo, procura-se expor como uma base de imagens foi construída.

Base de Dados	Pessoas	Imagens	Resolução	Poses
FERET [8]	1199	30000	256x384	20
CVL Face [10]	114	798	640x480	
PIE [5]	68	41368		13
BioID [1]	23	1521	384x286	
M2VTS [2]	295	2460	720x576	
FEI [3]	200	2800	640x480	

**Tabela 1. Comparativo entre bases de dados de imagens faciais.**

Nesta tabela pode-se observar a variedade de poses e o grande número de pessoas que participaram da base de imagens FERET [8], base mais conceituada e utilizada em estudos da atualidade para pesquisas de reconhecimento facial. Isso pode ser visto nos artigos científicos relacionados ao reconhecimento facial, como [11], [12] e [14]. Outras bases bastante conhecidas são mostradas na Tabela 1, com menor número de pessoas mas com um grande número total de imagens, como é o exemplo da PIE [5]. Outra base mostrada na tabela é a base brasileira da FEI [3], que conta com 14 imagens por pessoa totalizando 2800 imagens. Porém todas as bases estudadas possuem pequena resolução.

Para a organização do ambiente de captura de imagens foi utilizado como base o experimento de Rubinfeld [13], que mostra os principais detalhes sobre a captura de imagens utilizada por ele. Este estudo ainda mostra alguns experimentos com câmeras diferentes, formas diferentes de iluminação, mas sempre tentando manter um padrão entre as imagens capturadas.

A normalização de imagens é tratada, principalmente, no artigo da NIST [9], onde também são definidos os padrões que foram adotados para a padronização de imagens faciais. Os autores de [3] também tratam da captura e alinhamento das imagens. Neste trabalho, o alinhamento das imagens foi feito manualmente.

## 3. Normas para Captura e Normalização de Imagens

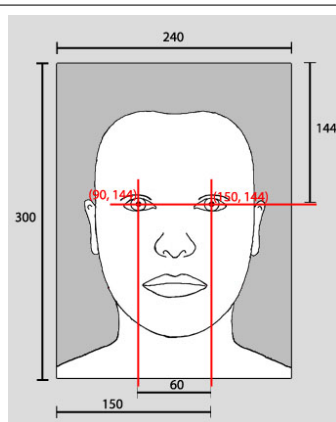
Com o objetivo de criar um banco de imagens para projetos de pesquisa na área de reconhecimento facial, as imagens foram capturadas de acordo com algumas normas, estabelecidas pelos documentos [7] e [9] e pelo relatório técnico M1/03-0114 [4].

Durante a captura da imagem facial, a pessoa deve olhar diretamente para câmera, sem que tenha qualquer tipo de inclinação do ombro ou corpo. Se existir qualquer inclinação, deve ser inferior a +/- 5° em relação à posição frontal, ou seja, tanto no lado esquerdo quanto no lado direito. A imagem deve mostrar claramente os olhos, principalmente a íris e a pupila. A presença de cabelos longos pode prejudicar as imagens obstruindo a parte da face. Isso deve ser evitado prendendo o cabelo ou colocando atrás das orelhas. Para mostrar claramente os olhos também é aconselhável não utilizar óculos [7].

Além da imagem frontal, outras imagens devem ser capturadas segundo Islam [7]: imagens das laterais direita e esquerda da face com rotação de 45° e 90°, e imagens frontais da face com inclinação entre 20° e 45° para baixo e para cima.

Outras normas importantes para o processo de captura de imagens são em relação ao equipamento a ser utilizado, à iluminação e ao tipo de fundo da imagem. Uma condição de iluminação adequada deve ser empregada para proporcionar uma iluminação uniforme à imagem que está sendo capturada e ao fundo. A iluminação deve ser distribuída igualmente em todos os lados da imagem sem gerar sombra e o local de captura deve estar totalmente isolado de iluminação natural (iluminação externa) para que não comprometa o experimento. O fundo da imagem deve ser limpo e sem objetos adicionais, para que, depois de capturada, a imagem mostre claramente as bordas da face em relação ao fundo. O fundo deve ter 18% de cinza e deve ser posicionado a aproximadamente 70 centímetros da pessoa.

O passo mais importante para a normalização de imagens é o posicionamento dos olhos. Através de dois pontos (pontos centrais das duas pupilas) é possível realizar grande parte da normalização. Segundo as normas, os olhos devem estar alinhados na mesma altura, para isso é utilizada a rotação da imagem. Em seguida, segundo o documento da NIST [9], a distância entre os olhos deve ser de 60 pixels, ou seja, é necessário realizar o redimensionamento da imagem. A imagem normalizada deve ter a resolução na proporção de 1:1.25 (ou 4:5) [7], com isso é necessário recortar a imagem para ficar com a resolução recomendada. A face deve ser posicionada na imagem de acordo com as medidas mostradas na figura (Figura 1) abaixo.



**Figura 1. Dimensões de uma imagem normalizada canônica (em pixels).**

Na Figura 1 pode-se notar que a distância entre os olhos é de 60 pixels e o tamanho total da imagem obedece a proporção 1:1.25, 240x300. Além disso o posicionamento da face é centralizada na imagem com os olhos localizados nos pontos (91, 144) e (150, 144). A Figura 1 representa uma imagem normalizada canônica, se for preciso uma resolução maior a imagem normalizada deve seguir as proporções indicadas das medidas desta imagem.

## 4. Metodologia

Nas subseções seguintes são mostrados os procedimentos utilizados para a captura e normalização de imagens neste trabalho.

### 4.1. Captura de Imagens

**4.1.1. Equipamentos** Para a captura de imagens foi utilizada uma câmera digital posicionada sobre um tripé para

que a distância entre a câmera e as pessoas não fosse alterada. A resolução de cada foto é de 2592x1944 pixels. Esta resolução foi escolhida por ser maior do que as existentes na literatura. E também permite ajustes posteriores quanto ao tamanho e à resolução, ou seja, será possível processar a imagem capturada para obter um tamanho menor, sem perder informações.

A câmera utilizada foi uma Sony Cyber-Shot (DSC-S3000) de 10.1 Mega Pixels. Porém, as imagens foram capturadas com 5 Mega Pixels. Para todas as posições de face a câmera estava com um valor de ampliação de três vezes (*zoom* de 3x). A ampliação é necessária para obter somente a face da pessoa, mas, ao mesmo tempo, a câmera deve ser posicionada numa distância da face para obter uma melhor nitidez e para evitar o aparecimento de sombras.

Um plano de fundo de cor fixa foi utilizado para a captura. A cor do fundo é cinza, sendo que a porcentagem aconselhada para captura de imagens faciais é de 18%, segundo Rubinfeld e Wilson [13].

A iluminação do ambiente de captura das imagens foi feita artificialmente, algumas imagens com fonte de iluminação própria para captura de imagens e outras com a iluminação da sala, caracterizando dois ambientes de captura de imagens: ambiente controlado e ambiente não controlado. Além dos ambientes controlado e não controlado, foram capturadas imagens com iluminação lateral para gerar sombra somente em um lado da face. O local de captura foi isolado de iluminação externa para não comprometer o experimento.

**4.1.2. Posicionamento do Equipamento** Com base nos estudos de Gross [6] e Rubinfeld [13] os equipamentos devem ser posicionados como mostrado na Figura 2.

Primeiramente, a cadeira foi posicionada um pouco a frente do painel de fundo para evitar sombras. Os pés da frente da cadeira estão a 1,2 metros do painel, fazendo com que a pessoa fique a aproximadamente 70 centímetros do fundo.

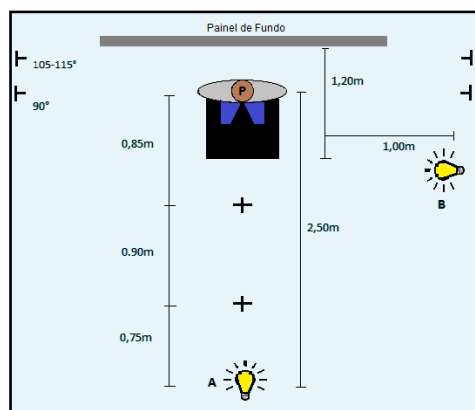
A câmera foi posicionada a 0,85 metros do centro da cadeira, de modo que somente a face e uma parte do ombro da pessoa apareçam na imagem adquirida. Posteriormente a câmera foi colocada a 1,75 metros do fundo da cadeira, para algumas fotos em escala. A variação das distâncias foi escolhida para ter uma maior variabilidade de imagens que podem ser utilizadas nos estudos de reconhecimento facial.

Com o objetivo de capturar as imagens com variações de iluminação, utilizou-se duas fontes de iluminação próprias, além da própria iluminação do ambiente da sala. Uma das fontes de iluminação foi posicionada a 2,5 metros do fundo da cadeira, de forma centralizada com a face. Esta posição evitou que aparecessem sombras no painel. A outra fonte foi localizada na lateral da cadeira, a 1 metro de distância e 1,7 metros de altura. Quando a fonte está ligada causa som-

bra em um lado da face. Neste estudo, isto serve para causar uma variação significativa de iluminação na imagem.

Pontos fixos foram definidos no chão para que o tripé/câmera, cadeira e iluminação sejam posicionados sempre no mesmo lugar. Isto é feito para que a distância entre a câmera e a face das pessoas seja sempre a mesma e a iluminação não sofra alterações. Além destes pontos, mais dois foram definidos para as fotos laterais (90° e 105-115°). De fato, estes pontos servem para auxiliar na captura da lateral da face. Para as fotos com a face voltada para cima foi definido um ponto junto à luz artificial para que as pessoas olhassem sempre para o mesmo ponto. Para as fotos com a face voltada para baixo outro ponto foi definido em um dos pés do tripé.

As posições do tripé assim como o posicionamento dos outros equipamentos podem ser melhor visualizados na Figura 2.



**Figura 2. Posicionamento do equipamento.**

Na Figura 2 os símbolos marcados com 'A' e 'B' são as iluminações próprias para a captura de imagens. A letra 'P' representa o local onde fica a pessoa. As marcações com o sinal de '+' são as duas posições do tripé e as marcas laterais (90° e 105-115°) são marcações para o voluntário olhar durante a captura das imagens laterais.

**4.1.3. Poses das Faces** As imagens foram capturadas da seguinte forma: imagens com iluminação controlada com a fonte de iluminação 'A' acesa e a iluminação ambiente desligada; imagens com iluminação não controlada com as fontes de iluminação 'A' e 'B' desligadas e a iluminação do ambiente ligada.

Durante a captura, é importante verificar que não existe cabelo cobrindo partes importantes como sobrancelhas e olhos e, principalmente, nas fotos laterais, a orelha deve aparecer descoberta.

O número total de imagens capturadas por pessoa neste estudo foi de 27, contando com todas as variações de iluminação e de pose. De modo geral, as imagens capturadas podem ser classificadas em duas categorias: 14 imagens com iluminação controlada e 13 imagens com iluminação não controlada. Foram capturadas imagens frontais com inclinações, imagens laterais e outras imagens com a câmera em posições diferentes. As poses das imagens capturadas são mostradas na Tabela 2.

**4.1.4. Procedimento para Captura das Imagens** Antes de iniciar a captura das imagens, num primeiro momento, os participantes devem assinar um termo de consentimento, dizendo que concordam e autorizam o uso de suas imagens para fins científicos.

Na Tabela 2 pode-se observar a codificação utilizada para identificar a pose de cada foto. Estes códigos são precedidos por outros números que representam o tipo de iluminação e a pessoa, de modo que a identidade possa ser preservada. Os códigos estão na sequência em que as imagens foram capturadas.

Código	Posição
01	90° Esquerda
02	105° Esquerda
03	Frontal
04	Frontal com Expressão
05	Inclinação Direita
06	Inclinação Esquerda
07	Face voltada para Cima
08	Face voltada para Baixo
09	90° Direita
10	105° Direita
11	2,5m Frontal
12	2,5m Inclinação Direita
13	2,5m Inclinação Esquerda
14	Frontal com luz lateral

**Tabela 2. Sequência de captura das imagens.**

A captura é realizada em duas etapas, uma com iluminação artificial ou controlada e outra com iluminação do ambiente ou não controlada. Apenas a imagem 14 (Tabela 2) é capturada somente com iluminação controlada utilizando, também, a fonte lateral de iluminação. Nas imagens 01 e 02, captura-se a face lateral direita do voluntário em 90 graus e com uma leve inclinação para direita em 105 a 115 graus. O mesmo é feito para as imagens 09 e 10, desta vez com a face lateral esquerda. As imagens 03 até 06 são frontais, com variações de inclinação e expressão facial. A imagem 07 é capturada com a pessoa olhando para cima e a 08 olhando para baixo. As ima-

gens 11, 12 e 13 são semelhantes às imagens 03, 05 e 06, respectivamente, porém com a câmera com uma distância de 1,75 metros. A última, número 14, é capturada com a pessoa olhando para a câmera e com uma luz lateral gerando sombra em um dos lados da face.

Contando com a utilização de iluminação artificial o flash da câmera deve estar desativado e o tempo do obturador da máquina, tempo de exposição, *zoom*, entre outras características da câmera devem ser fixos. Com isso, será possível ter um controle maior sobre as imagens capturadas. Este controle é necessário para as devidas comparações com outras bases de dados ou entre algoritmos e para validação da base. Outra forma de validação é a utilização de rotações e iluminação fixas. Porém são capturadas imagens com rotações diferentes e variações de iluminação para se aproximar da realidade, onde as imagens são capturadas sem controle.

#### 4.2. Normalização das Imagens

Nas subseções seguintes são apresentados os procedimentos utilizados para a normalização das imagens capturadas. O algoritmo de normalização foi implementado em C/C++ utilizando a biblioteca OpenCV<sup>1</sup>. Neste trabalho não foram implementadas a normalização de iluminação e fundo. E o centro dos olhos são obtidos manualmente.

Como especificado em [7] e [9] as imagens podem ter o tamanho de 768 X 1024 pixels mas, como já foi dito, as imagens neste trabalho foram capturadas com uma resolução maior. Porém, as novas medidas devem obedecer à proporção com as medidas das normas. As imagens devem ser normalizadas no tamanho da imagem canônica (Figura 1) ou com as medidas proporcionais, se desejável uma resolução maior. Neste estudo as imagens foram normalizadas com as medidas cinco vezes maior que a imagem canônica, logo a distancia entre os olhos é de 300 pixels, sendo que os olhos devem estar nas coordenadas (450, 720) e (750, 720) e o tamanho da imagem é de 1200x1500 pixels.

**4.2.1. Rotação** Esta etapa consta em alinhar os dois olhos da face. Para isso calculou-se a diferença de pixels entre os dois olhos. Primeiramente a diferença entre os valores x dos pontos e em seguida os valores y, representando os catetos de um triângulo. Com os catetos é possível calcular a hipotenusa e com esses valores é possível calcular o ângulo entre os olhos através da equação trigonométrica do arco cosseno.

**4.2.2. Redimensionamento** Para que existam 300 pixels entre os olhos é necessário redimensionar a imagem. Neste

passo é feita a divisão da distância entre os olhos por 300. Ou seja, como tem-se o valor da hipotenusa basta dividi-lo por 300 obtendo, assim, o valor de escala necessário para causar a ampliação ou redução da imagem.

**4.2.3. Centralização e Recorte da Face** Através do valor da escala da etapa anterior é possível calcular onde os pontos centrais estão localizados na nova imagem. Com isso é possível calcular o ponto central dos olhos para centralizar a face na imagem. Com isso basta cortar a imagem nos redores da face, mantendo ela centralizada e obedecendo as medidas citadas anteriormente. A imagem final, normalizada, deve ter o tamanho de 1200x1500 pixels.

### 5. Resultados

Na Figura 3 é mostrado o resultado da captura das 27 imagens de um voluntário. As imagens são mostradas na sequência determinada pela Tabela 2, sendo as 14 primeiras com iluminação controlada e as outras 13 com iluminação não controlada.

Encontrou-se dificuldade neste trabalho em relação ao posicionamento dos equipamentos, principalmente, para eliminar qualquer sombra na imagem. Este problema envolvia distância entre a câmera e a pessoa, entre a pessoa e o painel de fundo e altura da a fonte de iluminação.

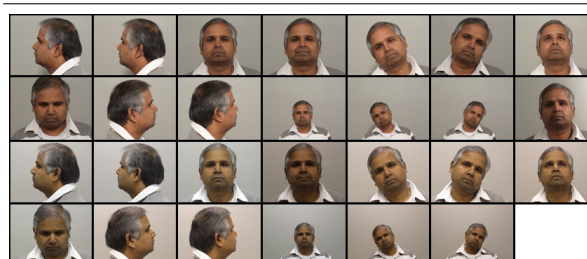


Figura 3. Resultado da captura de imagens.

A Tabela 3 mostra um resumo das características da base criada. Dentre o total de imagens que foram capturadas atualmente 65% são de pessoas do sexo masculino e 35% do sexo feminino. Quase 54% da base de imagens é formada por pessoas de faixa etária entre 21 e 30 anos. As imagens da base contém três variações de iluminação, levando em consideração a iluminação lateral, e um total de 14 poses diferentes para cada voluntário.

As imagens frontais passaram por um processo de validação. Este processo consta em verificar se elas estão dentro dos limites estabelecidos pelas normas citadas anteriormente. Esta validação utiliza o algoritmo de normalização de faces, sendo assim, as imagens foram validadas e a normalização foi testada. A tolerância

<sup>1</sup> *Open Source Computer Vision*. Disponível em: <http://opencv.willowgarage.com/wiki/>.

Resolução	2592x1944
Número de Pessoas	150
Número de Poses	14
Formas de Iluminação	3
Imagens por Pessoa	27
Total de Imagens	4050

**Tabela 3. Características da Base.**

de inclinação para os lados é de 5° e, em aproximadamente 9% das imagens esse valor foi maior. Dentre as imagens dentro do permitido a maior parte estava com inclinação abaixo de 3°. A maior inclinação das imagens frontais capturadas foi de 8,5° e a menor inclinação foi de 0,6°. Com isso é possível notar que a captura de imagens foi eficiente para grande parte das faces.

Seguindo o padrão das normas e através das equações citadas anteriormente as imagens da base de faces foram normalizadas. Vários testes foram feitos com as imagens rotacionadas para melhor visualização do processo de normalização. Na Figura 4 são mostrados dois exemplos deste processo. A primeira e a segunda imagem rotacionadas, sendo que a segunda também está em escala menor.



**Figura 4. Resultado da normalização de imagens.**

Nesta figura é possível ver a normalização passo a passo: na primeira coluna estão as imagens originais; na segunda coluna, as imagens rotacionadas com os olhos na mesma linha; na terceira, o redimensionamento da imagem para a diferença entre os olhos ficar em 300 pixels; e por último a imagem recortada para o tamanho de 1200x1500. Neste exemplo, a rotação para as imagens foi de 22° e 18°, respectivamente, e o *zoom* necessário para redimensionar as imagens para o tamanho correto foi de 0,82 e 1,57.

## 6. Conclusões

O objetivo deste artigo foi mostrar uma metodologia de captura de imagens faciais para a construção de uma base de imagens. Este estudo também apresenta a adoção de normas

e orientações e, ainda, reforça a necessidade de diversos cuidados para capturar as imagens. As imagens estão sendo utilizadas em estudos de normalização, detecção e reconhecimento facial. A principal dificuldade neste trabalho foi em relação ao posicionamento dos pontos de iluminação para eliminar a sombra nas imagens. Outro ponto que exigiu um esforço adicional foi em encontrar voluntários para colaborar na captura de imagens. O processo de normalização deste trabalho não tratou da normalização da iluminação nem do fundo de imagem. Isso será tratado em trabalhos futuros. Embora a base tenha alcançado, até o momento, um número razoável de imagens, o processo de captura ainda continua na tentativa de obter uma proporção equilibrada nas imagens em algumas categorias. Com este estudo percebe-se que a criação de um banco de imagens de alta qualidade é um procedimento complexo e criterioso.

A continuação deste trabalho tem como objetivo a detecção automática dos olhos. Com isso todo o processo de normalização passa a ser automático, facilitando o estudo das demais técnicas de reconhecimento facial.

## Referências

- [1] BioID. Bioid face database, 2011.
- [2] C. H. Cham. The extended m2vts database, 2006.
- [3] L. L. de Oliveira Junior and C. E. Thomaz. Captura e alinhamento de imagens: Um banco de faces brasileiro. Technical report, Centro Universitário da FEI, 2005.
- [4] P. Griffin. Face recognition format for data interchange. Technical report, Identix Corporate Research Center, 2003.
- [5] R. Gross. Pie database, 2000.
- [6] R. Gross. Face databases. *Handbook of Face Recognition*, 2005.
- [7] R. Islam. *Police Standard for Still Digital Image Capture and Data Interchange of Facial/Mugshot and Scar, Mark and Tattoo Images*. National Policing Improvement Agency, 2.0 edition, Maio 2007.
- [8] NIST. The feret database, 2004.
- [9] NIST. *Face Recognition Format for Data Interchange - Best Practices*, 2007.
- [10] P. Peer. Cvl face database, 1999.
- [11] P. J. Philips and S. A. Rizvi. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1090 – 1104, 2000.
- [12] S. A. Rizvi, P. J. Philips, and H. Moon. The feret verification testing protocol for face recognition algorithms. Technical report, NIST, 1998.
- [13] M. Rubinfeld and C. Wilson. Gray calibration of digital cameras to meet nist mugshot best practice. 1999.
- [14] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:775 – 779, 1997.

# Eye detection Approaches based on AdaBoost and SVM algorithms

Flávio das Chagas Prodossimo  
Universidade Tecnológica Federal  
do Paraná  
Curitiba, Paraná  
Email: frovas@gmail.com

Chidambaram Chidambaram  
Universidade Estadual  
de Santa Catarina  
São Bento do Sul, Santa Catarina  
Email: chidambr@gmail.com

Heitor Silvério Lopes  
Universidade Tecnológica Federal  
do Paraná  
Curitiba, Paraná  
Email: heitorslopes@gmail.com

**Resumo**—Face Recognition is an area of Computer Vision which is widely studied today. Face Recognition depends on many steps for its implementation, one of this steps is the detection of eyes. This can be used to extract features or as an initial step for the normalization of face images. In this work two approaches are proposed to detect the eyes. First approach is based on the Viola and Jones and the second approach is based on LBP (Local Binary Pattern), VNS (Variable Neighborhood Search) and SVM (Support Vector Machines). To evaluate both approaches, computational experiments are done. According to the detection performance, the Viola and Jones overcomes the SVM. However, both approaches are effective in detecting eye regions.

## I. INTRODUÇÃO

O reconhecimento facial (RF) está sendo aplicado em diversas áreas, como exemplo, no combate à falsificação de passaportes, identificação de criminosos e acesso a áreas restritas. Com os avanços e o fácil acesso a tecnologias, a captura de imagens por câmeras de segurança se tornou comum e vem aumentando a quantidade de imagens e vídeos. Além disso, nos últimos anos, vem crescendo o desenvolvimento de métodos e algoritmos nesta área. Neste contexto, a demanda por sistemas automáticos de RF cresceu nos últimos anos. A comunidade científica tem feito várias pesquisas com o objetivo de identificar pessoas usando somente a imagem da face. Para tanto, empregam-se técnicas e métodos científicos de diversas áreas, tais como Visão Computacional, Reconhecimento de Padrões, Inteligência Computacional e Métodos de Busca [20].

De modo geral, o reconhecimento de pessoas através de faces ainda constitui uma tarefa desafiadora, pois as imagens que contém faces possuem alta variabilidade em tamanho, forma, cor e textura. Portanto, nas últimas décadas, inúmeras técnicas vêm sendo desenvolvidas para identificar faces. Desta forma, cresce a demanda por métodos eficientes e robustos para a localização de faces humanas em imagens ou vídeos digitais.

O RF abrange diversas etapas da Visão Computacional. Estas etapas vão desde a captura de imagens para treinamento do sistema até a fase de reconhecimento. As etapas incluem detecção da face, detecção dos olhos, normalização da face, extração de características e a comparação de base de imagens com a imagem de busca [15], [20]. Dentro deste cenário, uma etapa importante é a Detecção de Regiões de Interesse (*ROI - Region of Interest*). Estas regiões podem ser os olhos, a face ou outras regiões que forneçam algum tipo de informação útil

para o reconhecimento, no caso deste trabalho, a região dos olhos ou os olhos em si. Com isso é possível realizar a etapa de normalização, que é responsável por padronizar a posição e o tamanho de elementos da face em relação à imagem.

O principal objetivo deste trabalho é detectar o centro dos olhos para normalizar as imagens faciais. Consequentemente, a detecção de regiões dos olhos entre vários componentes faciais se torna prioritária pois os olhos são os elementos faciais mais relevantes para o reconhecimento. O posicionamento dos olhos é utilizado em outras etapas, na extração de características, por exemplo [20]. Portanto, o foco principal deste trabalho é comparar duas metodologias de detecção de ROIs. A primeira baseada no algoritmo de Viola e Jones [18], o algoritmo clássico mais utilizado na literatura, principalmente para detecção de faces em imagens com várias pessoas. Na segunda abordagem utiliza-se um algoritmo de extração de textura LBP (*Local Binary Pattern*) [12] em conjunto com a Máquina de Vetores de Suporte, SVM (*Support Vector Machine*) [17] e o algoritmo de busca, VNS (*Variable Neighborhood Search*) que é uma Meta-heurística [10].

Diversos algoritmos podem ser utilizados para auxiliar a Detecção de ROIs. A seguir são apresentados alguns filtros de imagens, além de algoritmos de reconhecimento de padrões de otimização de busca, que visam facilitar as metodologias de detecção.

As próximas sub-seções dão um embasamento teórico sobre o trabalho desenvolvido e também apresentam-se alguns trabalhos relacionados (Seção I-F). Na Seção II está a explicação sobre o desenvolvimento deste trabalho. Alguns experimentos e os resultados são apresentados na Seção III. E, por fim, na Seção IV, são expostas as considerações finais.

### A. Local Binary Pattern

O filtro de texturas, LBP foi proposto por [12]. A ideia principal deste filtro foi baseado em um modelo de análise de texturas, onde a região de textura era representada por 8 elementos (uma região de vizinhança de 3x3 pixels). Neste modelo, cada um dos elementos pode receber 3 valores (0, 1 e 2), totalizando em 6561 ( $3^8$ ) texturas diferentes. Já na versão atual, cada elemento pode ser dividido em 2 valores, com isso reduz o número de possíveis texturas para 256. A Figura 1 representa o processo de aplicação do filtro LBP.

O exemplo da Figura 1 mostra que a imagem original (a) é analisada a partir do pixel central. Todos os pixels vizinhos



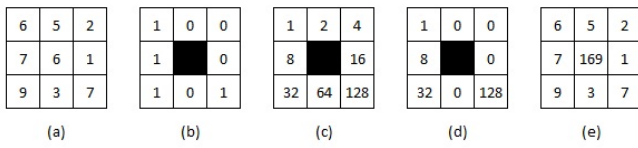


Figura 1. Procedimento do LBP [12].

são analisados e recebem valor 1 caso seja maior ou igual ao pixel central e 0 caso contrário (b). Os valores binários são multiplicados, elemento a elemento, pela matriz de valores na potência de 2 (c), resultando na matriz (d). Por fim o pixel central recebe a soma dos resultados da matriz (d), resultando na imagem (e).

O LBP é um filtro geralmente utilizado para classificação, por isso é necessário um algoritmo de Reconhecimento de Padrões para treinar e classificar as imagens processadas pelo filtro. O algoritmo mais utilizado na literatura em conjunto com o LBP é o SVM, explicado na Seção I-C.

**B. Algoritmo Viola e Jones**

Esta abordagem de aprendizado de máquina para detecção de objetos foi proposto por Viola e Jones [18]. Suas três principais contribuições marcantes são: a representação de imagens através de imagens integrais para extrair características de modo rápido em várias escalas, onde as imagens foram filtradas com o filtro *Haar-like Features*; a construção de um classificador selecionando um conjunto pequeno de características através do Adaboost; e por último, a construção de um classificador complexo numa estrutura cascata que drasticamente aumenta a velocidade do detector focando apenas nas regiões promissoras. Neste trabalho, as características *Haar-like* foram utilizadas na detecção por conterem informações mais relevantes do que uma pequena quantidade de pixels individualmente.

A Figura 2 mostra o processo de execução do Viola e Jones, onde em cada quadro as características são extraídas para fazer o reconhecimento com o Adaboost. As características são calculadas da seguinte forma: soma-se os níveis de cinza da região destacada em preto no retângulo e subtrai da soma dos níveis de cinza da região branca do retângulo.



Figura 2. Exemplo de execução do algoritmo Viola e Jones.

**C. Máquinas de Vetores de Suporte - SVM**

A classificação através de algoritmos de reconhecimento de padrões, por exemplo, SVM, é feita através de treinamento, que consiste na apresentação de diversos dados para o algoritmo que, por sua vez, estabelecerá um padrão. As etapas básicas de um algoritmo de reconhecimento de padrões podem ser vistos na Figura 3. A primeira etapa consiste na obtenção de vários dados de exemplo que serão utilizados pelo algoritmo no treinamento ou teste. Estes dados passam pelo Pré-Processamento, e então, extrai-se as características relevantes para a classificação. Na etapa seguinte ocorre o treinamento ou teste usando o algoritmo de classificação.

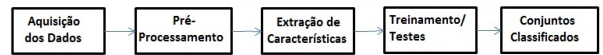


Figura 3. Etapas de um algoritmo de Reconhecimento de Padrões.

As SVMs são caracterizadas pelo aprendizado estatístico que consiste em definir uma série de pontos em um espaço multidimensional e encontrar um hiperplano que separe esses pontos em classes diferentes. Isto é, o algoritmo estabelece relações matemáticas para criação de um classificador de acordo com o conjunto de dados que lhe foi apresentado para treinamento. Os dados treinados pelo SVM clássico podem ser divididos em somente duas classes distintas.

**D. Variable Neighborhood Search (VNS)**

Os algoritmos de busca são responsáveis por encontrar a melhor solução possível dentro de um espaço de busca através de tentativas. Um exemplo deste tipo algoritmo é chamado VNS, proposto por [10]. Esta tem como princípio básico pesquisar no espaço de busca utilizando o princípio da vizinhança. O VNS prossegue a busca através de método descida, para um local mínimo e, então, explora uma série de vizinhanças pré-determinadas. A partir destes locais mínimos, tenta escapar dos vales com a perturbação. Ao contrário de outras Meta-heurísticas, o VNS não segue nenhuma trajetória, porém, explora cada vez mais as vizinhanças distantes em relação à solução atual. Atua numa vizinhança e muda para uma nova somente se encontrar uma melhoria em relação à atual vizinhança. Além do mais, numa determinada vizinhança, a rotina da busca local é repetidamente aplicada para encontrar um ótimo local.

A estrutura de vizinhança pode ser definida como um conjunto finito pré-selecionado de estruturas de vizinhanças, representado por  $N_k$ , onde  $k = 1, \dots, k_{max}$  é o conjunto de soluções do k-ésimo vizinho de  $x$ , por  $N_k(x)$ . O pseudocódigo básico do VNS é apresentado abaixo [5]:

*Inicialização:*

- Selecione um conjunto de estruturas de vizinhança  $N_k$ ;
- Encontre uma solução inicial  $x$ ;
- Defina uma condição de parada;
- Repita até atender o critério de parada:
  - i.  $k = 1$ ;
  - ii. Repita até  $k \leftarrow k_{max}$ :
    - Perturbação de variáveis; (Gera um ponto  $x'$  aleatoriamente do k-ésimo vizinho de  $x$  onde  $x' \in N_k(x)$ );

- Busca local usando como a solução inicial  $x'$  selecionando  $x''$  como ótimo local;
- Mudança de vizinhança:  
Se  $x''$  for melhor que o inicial ( $x$ ), então mude para a nova vizinhança ( $x \leftarrow x''$  e continua busca  $N_1(k \leftarrow 1)$ , caso contrário,  $k \leftarrow k + 1$  ;

O critério de parada pode ser determinado por um número máximo de iterações ou número máximo de iterações entre duas melhorias [5]. A vantagem do VNS é que este permite encontrar soluções ótimas ou próximas à ótima em um tempo moderado. O VNS é simples, rápido e exige apenas um parâmetro que é o número de estruturas de vizinhança. Além do mais, esta meta-heurística pode ser facilmente hibridizada [10]. Este algoritmo foi inicialmente focado nas aplicações de otimização combinatorial e também, nos problemas de otimização global.

#### E. Detecção de Regiões de Interesse

A Detecção de ROIs é importante para o RF, pois a partir destas regiões é que se iniciam outros processos, como a normalização e a extração de características. Estas regiões geralmente incluem a face, a região dos olhos (Figura 4) e o centro do olho (pupila). Para a etapa de normalização de faces, o mais importante é a pupila onde o ponto central de cada um dos olhos deve ser detectado.

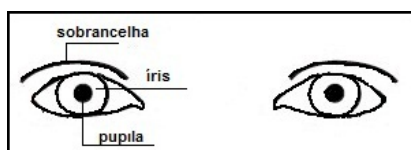


Figura 4. Região dos Olhos.

os métodos de detecção de olhos são classificados em ativo e passivo. O método ativo é usado para detectar olhos em sequências de vídeos enquanto o método passivo é mais utilizado em imagens. Este método possui três abordagens: análise de aparência, aprendizagem e modelos [2]. Na primeira abordagem, análise de aparência, são utilizadas características de cor, textura e formas. Agrupamentos de pele são formados em algum espaço de cor, feitos através de análise estatística de pele humana presente nas imagens. As vantagens desta abordagem são a necessidade de poucos recursos computacionais e, também, pelo fato de que os cálculos de características serem invariantes a escala, rotação, cores ou formas. A principal desvantagem é a dificuldade de estabelecer critérios para identificar pele de várias etnias e em imagens com variação de iluminação [2].

Para a abordagem de aprendizagem um classificador é treinado utilizando um conjunto de imagens de exemplos e de contraexemplos. O treinamento pode ser feito utilizando técnicas como: Redes Neurais, SVMs, entre outras. Nesta abordagem, os problemas de rotação, escala e iluminação podem ser anulados pela geração de vários exemplos que representem estas características. A maior desvantagem está na definição de um bom contraexemplo para o treinamento [2].

A detecção por modelos tem como o princípio básico o casamento de uma imagem modelo com regiões da imagem onde há face. Como exemplo, pode-se citar o *Template Matching* (TM). No processo por TM, geralmente, procura por um padrão em uma imagem paisagem (imagem com fundo, cenário, múltiplas faces etc.), ou seja, o melhor casamento entre uma imagem referência ou de busca e cada recorte (janela) da imagem paisagem. Uma imagem modelo ou de referência podem ser a de imagens de olhos recortadas de uma face ou um padrão composto por um círculo e duas curvas, formando um modelo de olho [2]. A principal desvantagem deste método é o alto custo computacional, pelo grande espaço de busca e pela necessidade de modificar o modelo em relação à escala e rotação.

#### F. Trabalhos Relacionados

Embora o foco do trabalho seja de detectar os olhos, o processo quase sempre se inicia com a detecção de faces. Os métodos existentes podem ser agrupados em três grandes áreas: Métodos baseado em conhecimento, Abordagens baseado em características invariantes e Métodos que utilizam *templates* [19]. Na literatura, é mais comum encontrar trabalhos relacionados às últimas duas áreas. No trabalho de [8], a detecção de olhos foi feita seguindo uma série de etapas que incluem a segmentação da pele, a detecção de bordas, as operações morfológicas, o cálculo de medidas geométricas entre os olhos, e por último, a aplicação de diversos *templates*. A melhor taxa de acerto foi de 91,23% para faces com rotação menor que 20° e o pior caso foi para rotações maiores que 60° onde a taxa de acerto foi de 62,21%.

De modo similar ao trabalho anterior, em [9] também foi utilizada a detecção da região da face através da cor da pele utilizando-se o espaço de cores YCrCb. Neste, a região de interesse dos olhos foram detectados aplicando-se o filtro de Gabor. Neste trabalho, Ke e Huang [9] obtiveram 94% de acerto com imagens frontais e 83% para imagens com rotação. No estudo de [7], utilizou-se imagens em níveis de cinza com baixa resolução. Neste trabalho, primeiramente, utilizando o algoritmo Viola e Jones [18], detectou-se a face, e então, foram detectadas duas regiões, uma para cada olho. Na etapa seguinte, a posição da pupila de cada olho foi localizada através do TM, utilizando a soma das diferenças absolutas. [7] perceberam que, quando a imagem da face é transformada para níveis de cinza, a região central do olho é formada por um círculo preto, e assim quatro *templates* sintetizados foram definidos. Utilizando esses *templates* formados por um círculo preto e uma pequena borda branca, a melhor taxa média de acerto, em torno de 95,5%, foi obtida.

No trabalho de [13] foi utilizada a técnica de TM. As imagens utilizadas nos experimentos não passaram por nenhum pré-processamento. Os *templates* utilizados nos experimentos foram obtidos das próprias imagens faciais. A busca para o *matching* dos *templates* foi otimizada com dois algoritmos de otimização baseada em inteligência de enxames, Colônia de Abelhas, ABC (*Artificial Bee Colony*), utilizado por [1], e Busca Harmônica, HS (*Harmony Search*), utilizado por [16]. Da mesma forma, o ABC também foi utilizado em estudos de detecção de íris por [7].

Para a detecção de sorrisos no trabalho de [3] foi utilizada uma combinação do LBP e SVM, a primeira para extração

de características e a segunda, classificação das imagens. Para o treinamento da SVM foram utilizadas imagens de 59x65 pixels, sendo que existiam 2421 imagens de faces com sorriso e 3360 sem sorriso. Nos melhores casos a taxa de acerto do algoritmo de [3] foi de aproximadamente 90%.

## II. DESENVOLVIMENTO

O desenvolvimento deste trabalho é baseado em duas abordagens de detecção de olhos que é essencial para normalização automática de faces. A primeira abordagem é baseada em algoritmo clássico de Viola e Jones e a segunda, é baseada no SVM que classifica as imagens. Porém, na segunda, foi utilizada LBP para extrair as características de textura e VNS, para realizar a busca no espaço de imagem levando em consideração a escala e a rotação. O algoritmo Viola e Jones foi adaptado para detectar outras regiões, não somente a face como o algoritmo original. Portanto, o algoritmo utilizado neste trabalho localiza a região de olhos numa imagem facial.

O procedimento de detecção dos olhos (Figura 5) utilizado foi o mesmo para as duas abordagens, diferindo somente pelo algoritmo de classificação e método de extração de características. No caso do algoritmo Viola e Jones, primeiramente a imagem é apresentada ao sistema que detecta uma região retangular em que estão situados os dois olhos. Em seguida, dentro desta região, o algoritmo busca os dois olhos, retornando ao mesmo tempo, o ponto central de cada um dos olhos. Para extração de características, o Viola e Jones utiliza *Haar-Like Features* e o treinamento é feito usando Adaboost.

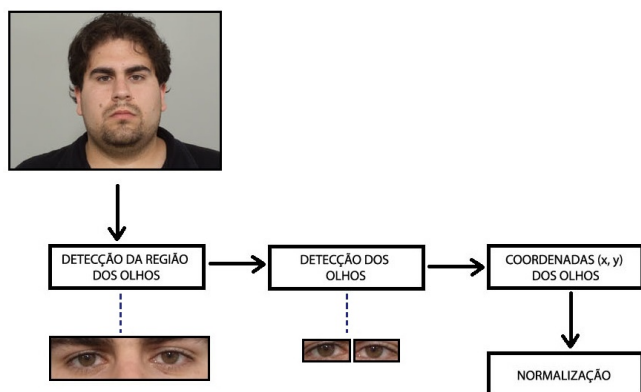


Figura 5. Procedimento de Detecção dos Olhos.

Na abordagem baseada em LBP e SVM, inicia-se a busca por uma região ótima na imagem facial que melhor represente uma região de olhos. Então usando os parâmetros gerados pelo VNS, a imagem retangular é recortada da imagem facial (paisagem) e pré-processada usando o filtro de textura LBP. Assim, o vetor de características da imagem é construído. Este vetor servirá como o exemplo de teste para o SVM que classificará a região cortada como uma região de olhos ou não. Assim, o processo de busca por uma região ótima será repetida através de várias iterações pré-determinadas. Para treinamento do algoritmo proposto foram utilizadas 483 imagens de regiões de olhos (Figura 5) e, em seguida, foram utilizadas 966 imagens contendo somente os olhos. Com isso pode-se realizar as duas etapas da detecção de forma separada.

Em seguida, foram realizadas validações do algoritmo utilizando o mesmo número de imagens do treinamento, ainda utilizando o algoritmo SVM. Com o algoritmo validado e funcionando, o SVM foi re-treinado juntando o conjunto de treinamento e o conjunto de validação, resultando em 966 imagens para re-treino dos recortes da região dos olhos e 1932 imagens para recortes dos olhos. O treinamento final foi armazenado em arquivos para serem utilizados nos testes, sem a necessidade de novos treinamentos a cada execução da detecção dos olhos.

## III. EXPERIMENTOS E RESULTADOS

Os experimentos foram realizados utilizando a base de imagens faciais apresentada no trabalho de [14]. Com o objetivo de criar um banco de imagens para projetos de pesquisa na área de reconhecimento facial, as imagens frontais foram capturadas de acordo com algumas normas, estabelecidas pelos documentos [6] e [11] e pelo relatório técnico M1/03-0114 [4]. A iluminação do ambiente de captura das imagens foi feita artificialmente, algumas imagens com fonte de iluminação própria e outras com a iluminação da sala, caracterizando dois ambientes. A base de imagens contém 138 classes de indivíduos diferentes. A base foi criada em 2012 com intuito de ser pública, porém, ela ainda está em fase organização e testes. Somente as imagens frontais foram utilizadas e estão agrupadas usando os códigos a seguir: FI1, FI1exp, FI1esc e FI1II são imagens frontais com a fonte de iluminação 1, sem expressão, com expressão, com escala e com sombra lateral, respectivamente; e os códigos FI2 e FI2exp são imagens frontais com a segunda forma de iluminação, sem expressão e com expressão, respectivamente. Exemplos das imagens, nesta mesma ordem, são apresentadas na Figura 6.

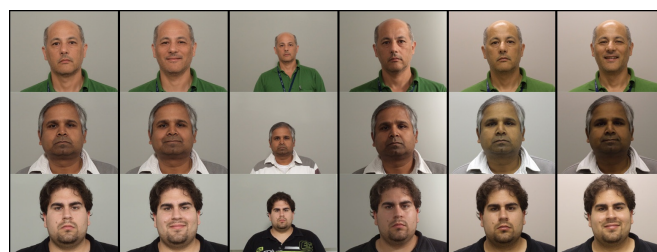


Figura 6. Exemplos das imagens utilizadas nos experimentos.

Com o objetivo de avaliar as duas abordagens, realizaram-se dois experimentos utilizando esta base de imagens. Os resultados são apresentados nas Tabelas I e II. Na primeira abordagem as soluções não variam quando o algoritmo é executado mais de uma vez. Já no outro algoritmo as soluções mudam, pelo fato de se utilizar um algoritmo baseado em meta-heurística de busca onde existe uma certa aleatoriedade para iniciar as variáveis. Por isso, na Tabela II os valores mostrados são o resultado da média entre várias execuções.

Com os resultados apresentados na Tabela I, é possível notar que o algoritmo de Viola e Jones supera o SVM tanto no número de acertos quanto no tempo de processamento para encontrar os olhos. O melhor número de acertos foi para as imagens frontais com a iluminação 2, em compensação foi o que mais demorou, em média, para encontrar os olhos na face. No geral, o melhor caso pode ser observado com os testes das

imagens com escala e iluminação 1, onde o tempo médio foi o menor dentre os testes, errando somente em duas faces.

Tabela I. RESULTADOS DA DETECÇÃO DE OLHOS COM O ALGORITMO VIOLA E JONES.

Pose	Tempo Médio (s)	Tempo Total (s)	Acertos
FI1	2,2	305	131
FI1exp	2,7	373	126
FI1esc	2,1	300	136
FI1II	2,7	385	134
FI2	2,9	404	137
FI2exp	2,2	304	135

A Tabela II mostra os resultados obtidos da média de 20 rodadas de execução do algoritmo proposto usando SVM e VNS, já o desvio padrão das execuções é mostrado no gráfico da Figura 7. Na Tabela II observa-se que este algoritmo é mais lento, porém, na maioria dos casos é eficiente para encontrar os olhos na face. Sendo que, na média, o maior número de acertos foi 131 com tempo médio por imagem de aproximadamente 27 segundos. Porém, nas imagens com a escala, a número de acertos foi de 108, bem abaixo do que encontrado por o algoritmo Viola e Jones que obteve 136 acertos.

Tabela II. RESULTADOS DA DETECÇÃO DE OLHOS COM O ALGORITMO LBP COM SVM.

Pose	Tempo Médio (s)	Tempo Total (s)	Acertos
FI1	26,5	3664	130
FI1exp	26,9	3725	125
FI1esc	26,9	3721	108
FI1II	22,7	3135	130
FI2	27,1	3753	131
FI2exp	27,5	3795	124

O gráfico da Figura 7 mostra os resultados de todos os testes, onde cada coluna representa uma categoria de imagens. Sendo que os teste do algoritmo Viola e Jones sempre apresenta o mesmo número de acertos, independente de número de execuções. Isso está representado por uma linha vermelha no gráfico. Como o VNS é o algoritmo de otimização, portanto, na segunda abordagem com SVM, os resultados são mostrados usando o desvio padrão (linha azul vertical) e a média (centro da linha azul) dos testes.

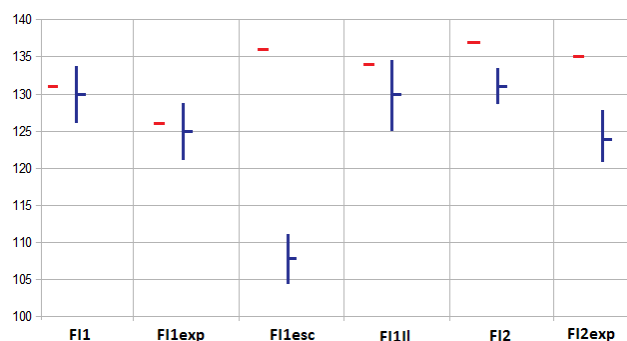


Figura 7. Comparação de número de acertos entre Viola e Jones e SVM

Conforme mostrado na Figura 8, a abordagem baseada em SVM e LBP atinge uma taxa de acertos aproximadamente menor que 26% a do Viola and Jones na categoria de imagens com escala e a iluminação própria (FI1esc). Ao mesmo tempo, na categoria de expressão facial com a iluminação ambiente

(FI2exp), a diferença de taxa de acertos entre as duas foi de 9%.

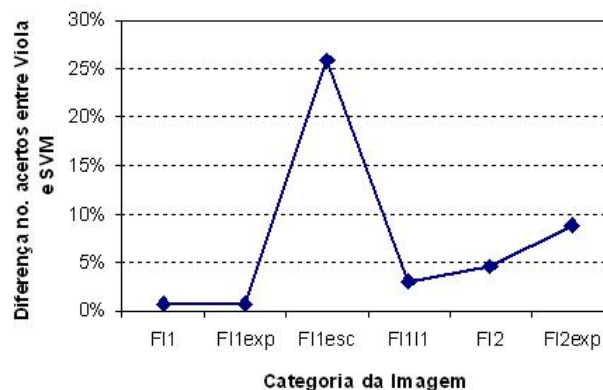


Figura 8. Diferença de número de acertos entre os algoritmos

Baseado na análise dos resultados, percebe-se que a região do nariz foi a que mostrou maior erro de detecção dentre todas as regiões. Isto pode ter acontecido pelo fato de existir uma semelhança de textura entre as regiões de olhos e de nariz caracterizada por dois círculos escuros dentro de uma região clara. Algumas imagens de regiões detectadas erroneamente podem ser observadas na Figura 9.



Figura 9. Regiões detectadas erroneamente.

#### IV. CONSIDERAÇÕES FINAIS

De acordo com os experimentos realizados, verifica-se que, de modo geral, as duas abordagens propostas são eficientes para detectar olhos em imagens de faces humanas. Porém, considerando o tempo de execução, o algoritmo de Viola e Jones se torna o mais rápido e robusto. Embora a abordagem baseada em SVM e LBP tenha alcançada resultados inferiores à Viola e Jones, a proposta poderá servir de referência para análises futuras e outros trabalhos relacionados. A continuação deste trabalho consta em normalizar as imagens da base mencionada [14] de forma automática, utilizando os pontos centrais dos olhos encontrados pelo algoritmo. Assim, construir uma base de imagens faciais normalizadas que é essencial no processo de reconhecimento de faces.

#### REFERÊNCIAS

- [1] C. Chidambaram and H. S. Lopes. An improved artificial bee colony algorithm for object recognition problem in complex digital images using template matching. *International Journal of Natural Computing Research*, 1(2):54–70, 2010.
- [2] B. de Brito Leite. Detecção de olhos em imagens com faces humanas. Master's thesis, Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2008.

- [3] D. Freire, M. C. Santana, and O. Déniz-Suárez. Smile detection using local binary patterns and support vector machines. *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications (VISAPP 2009)*, 1:5 – 8, 2009.
- [4] P. Griffin. Face recognition format for data interchange. Technical report, Identix Corporate Research Center, Minnetonka - MN, 2003. Relatório Técnico.
- [5] P. Hansen and N. Mladenovic. Variable neighborhood search: Principles and applications. *European Journal of Operational Research*, 130:449–467, 2001.
- [6] R. Islam. *Police Standard for Still Digital Image Capture and Data Interchange of Facial/Mugshot and Scar, Mark and Tattoo Images*. National Policing Improvement Agency, United Kingdom, 2.0 edition, Maio 2007. Manual.
- [7] J. C. S. J. Jacques, J. L. Moreira, A. Braun, and S. R. Musse. A template-matching based method to perform iris detection in real-time using synthetic templates. *Proceedings of IEEE International Symposium on Multimedia*, pages 142–147, 2009.
- [8] W. Jian and Z. Honglian. Eye detection based on multi-angle template matching. *Proceedings of International Conference on Image Analysis and Signal Processing*, pages 241 – 244, 2009.
- [9] L. Ke and Y. Huang. Eye location based on dual-orientation gabor filters and templates. *Proceedings of International Congress on Image and Signal Processing*, page 4, 2009.
- [10] N. Mladenovic and P. Hansen. Variable neighborhood search. *Computers and Operations Research*, 24:1097 – 1100, 1997.
- [11] NIST. *Face Recognition Format for Data Interchange - Best Practices*. United States, 2007. Manual de Boas Práticas.
- [12] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29:51 – 59, 1996.
- [13] F. C. Prodossimo, C. Chidambaram, and H. S. Lopes. Otimização da detecção de olhos em imagens faciais utilizando os algoritmos colônia de abelhas artificiais e harmony search. *Anais do X Congresso Brasileiro de Inteligência Computacional*, 2011.
- [14] F. C. Prodossimo, C. Chidambaram, and H. S. Lopes. Proposta de uma metodologia para a construção de um banco de imagens faciais normalizadas. *Anais do VIII Workshop de Visão Computacional*, 2012.
- [15] D. Reisfeld and Y. Yeshurun. Preprocessing of face images: Detection of features and pose normalization. *Computer Vision and Image Understanding*, 71:413–430, 1998.
- [16] M. H. Scalabrin, R. S. Parpinelli, and H. S. Lopes. Paralelização do algoritmo harmony search utilizando unidade de processamento gráfico. *Mecânica Computacional*, 29:7109–7121, 2010.
- [17] V. N. Vapnik. *The nature of Statistical learning theory*. Springer-Verlag, New York, 1995.
- [18] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:511 – 518, 2001.
- [19] M.-H. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 21(1):34–58, 2002.
- [20] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35:399–458, 2003.