



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I CAMPINA GRANDE
CENTRO CIÊNCIA E TECNOLOGIA
CURSO DE LICENCIATURA EM COMPUTAÇÃO**

DIELLY DE OLIVEIRA VIANA

Análise da Influência da Iluminação e da Oclusão na Classificação de Faces em Imagens Digitais

CAMPINA GRANDE – PB
2012

DIELLY DE OLIVEIRA VIANA

Análise da Influência da Iluminação e da Oclusão na Classificação de Faces em Imagens Digitais

Relatório apresentado ao Curso de Graduação em Licenciatura em Computação da Universidade Estadual da Paraíba, em cumprimento à exigência para obtenção do grau Licenciado em Computação.

Orientador: Eanes Torres Pereira

CAMPINA GRANDE – PB
2012

V614a Viana, Dielly de Oliveira.
Análise da influência da iluminação e da oclusão na classificação de faces em imagens digitais [manuscrito]. / Dielly de Oliveira Viana. – 2012.
51 f.: il. color.

Digitado
Trabalho de Conclusão de Curso (Graduação em Computação) – Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2012.
“Orientador: Prof. Dr. Eanes Torres Pereira, Departamento de Computação”.

1. Detecção de faces. 2. Invariância a iluminação. 3. Invariância á oclusão.
I. Título.

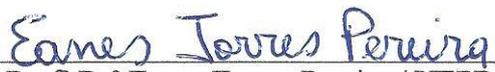
21. ed. CDD 006.42

DIELLY DE OLIVEIRA VIANA

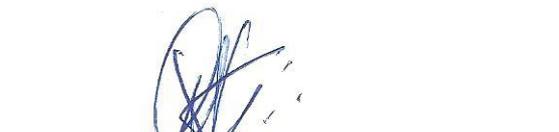
Análise da Influência da Iluminação e Oclusão na Classificação de Faces em Imagens Digitais

Relatório apresentado ao Curso de Graduação em Licenciatura em Computação da Universidade Estadual da Paraíba, em cumprimento à exigência para obtenção do grau Licenciado(a) em Computação.

Aprovado em 07/12/2012.


Prof.º Dr.º Eanes Torres Pereira / UEPB
Orientador


Prof. Dr. Robson Pequeno de Sousa / UEPB
Examinador


Prof. Me. Frederico Moreira Bublitz / UEPB
Examinador

AGRADECIMENTOS

Agradeço primeiramente a Deus por proporcionar momentos importantes como este em minha vida, pois sem Ele nada conseguiria.

Em seguida, aos meus pais Djaci Viana e Helena Alfa por estarem sempre ao meu lado com incentivo aos estudos. Além de todos os amigos e parentes, dentro e fora da universidade, que sempre me ajudaram e acreditaram no meu potencial.

A todos os professores do Curso de Licenciatura em Computação da UEPB, que sempre estiveram nos incentivando a conhecer novos horizontes e adquirir novos conhecimentos.

Um agradecimento especial ao Professor Doutor e Orientador deste trabalho, Eanes Torres Pereira, por sua dedicação e perfeição naquilo que faz de melhor, além de sua infinita paciência em todas as contribuições prestadas.

A todos os funcionários da UEPB, pela presteza e atendimento quando nos foi necessário.

APRESENTAÇÃO

O Trabalho de Conclusão de Curso tem por objetivo validar o período de aprendizado adquirido pelo aluno em sua graduação, avaliando-o conforme as exigências da Universidade Estadual da Paraíba para obtenção do grau no Curso de Licenciatura em Computação.

Com esses objetivos, foi visado o aproveitamento por parte do aluno, dos experimentos feitos em dois anos de iniciação científica no campo da visão computacional que gerou bastantes resultados significativos que não podiam ser desprezados pela sua importância.

Além disso, por esta área computacional está em rápido crescimento e com diversas universidades a oferecer continuidade nesta linha de pesquisa em programas de mestrado e doutorado, se fez totalmente válido o trabalho de conclusão de curso com o propósito de enriquecimento nesta área de conhecimento, o que pode abrir portas futuras.

No entanto, para a realização deste trabalho, foi necessária a aprendizagem e entendimento de diversos outros assuntos de conhecimento na área da visão computacional, o que deu larga experiência em diversos assuntos relacionados. As dificuldades encontradas nesta área ainda têm muito a serem exploradas, analisadas e estudadas.

RESUMO

Este trabalho consiste no estudo da dificuldade que os classificadores de padrões enfrentam na área da detecção de faces em imagens digitais quando testados em faces com condições de extrema iluminação em diversos ângulos de origem, o que conseqüentemente pode ocasionar um sombreamento na face. Como foi visto na literatura existente, a maioria dos experimentos realizados nesta área da detecção de faces se depara com problemas de iluminação, porém alguns dos autores propõem métodos para contornar o problema de forma inadequada. Sendo assim, veio a necessidade de avaliar o que já existe de mais eficiente na área e se alguma combinação de métodos de extração de característica junto com algum tipo de classificador contornaria essa necessidade. Foram analisados quatro métodos de extração de características: valores dos pixels da imagem, histogramas de cinza, análise de componente principal e padrões binários locais. Estes métodos foram combinados com dois tipos de classificadores: máquina de vetor de suporte e redes neurais artificiais. Os métodos foram extraídos das imagens de modo global e local de acordo com a característica de cada método. Para os testes foi usada uma base de imagens específicas, com variâncias no grau de iluminação. Após todos os experimentos realizados, foi aplicada a análise de variância entre os resultados, o que validou estatisticamente os experimentos. Com isso, foi possível ordenar em grau de importância quais os métodos mais eficazes para o problema de faces com iluminação. Dos métodos analisados, o Padrão Binário Local (LBP – Local Binary Patterns) dividindo a imagem em regiões com 3x3 pixels combinado com uma rede neural artificial como classificador obteve o melhor resultado geral.

PALAVRAS-CHAVE: Detecção de faces. Extração de característica. Invariância à Iluminação. Invariância à Oclusão. Classificadores.

LISTA DE TABELAS

Tabela 1 - Resultados dos experimentos globais.....	36
Tabela 2 - Resultados dos experimentos locais.....	37
Tabela 3 - Análise isolada dos resultados usando histograma de cinza.	38
Tabela 4 - Análise isolada dos resultados usando LBP.....	39
Tabela 5 - Análise isolada dos resultados obtidos em regiões com 7x7 pixels.....	40
Tabela 6 - Análise isolada dos resultados obtidos em regiões com 3x3 pixels.....	40
Tabela 7 - Distribuição dos valores convertidos usando <i>F-score</i> para análise de variância. ...	42
Tabela 8 - Resumo da análise geral da variância.....	43
Tabela 9 - Resultado da análise de variância ANOVA.....	43
Tabela 10 - <i>Ranking</i> dos métodos e classificadores pelas médias <i>F-score</i>	45

LISTA DE GRÁFICOS

Gráfico 1 – Representação gráfica de um histograma de níveis de cinza..... 27

LISTA DE FIGURAS

Figura 1 - Exemplo de uma rede neural artificial	24
Figura 2 - Mapeamento entre faixas de valores (bins) e posição dos pixels na imagem	26
Figura 3 - Exemplo do método LBP	29
Figura 4 - Quantidade de regiões e ordem de concatenação	33

LISTA DE SIGLAS

UEPB	Universidade Estadual da Paraíba
PCA	Análise de Componente Principal (<i>Principal Component Analysis</i>)
LBP	Padrão Binário Local (<i>Local Binary Patterns</i>)
INTLBP	<i>Integral Local Binary Pattern</i>
ANOVA	Análise de Variância
SVM	Máquina de Vetor de Suporte (<i>Support Vector Machine</i>)
RNA	Redes Neurais Artificiais
SQ	Soma dos Quadrados
GL	Grau de Liberdade
MQ	Média dos Quadrados

SUMÁRIO

1 - INTRODUÇÃO	12
2 - REVISÃO BIBLIOGRÁFICA.....	16
3 - FUNDAMENTOS TEÓRICOS	21
3.1 - VISÃO COMPUTACIONAL	21
3.2 - APRENDIZAGEM DE MÁQUINA.....	21
3.2.1 <i>SUPPORT VECTOR MACHINES</i> (SVM).....	22
3.2.2 REDES NEURAIS ARTIFICIAIS (RNA).....	23
3.3 - PROCESSAMENTO DE IMAGENS: EXTRAÇÃO DE CARACTERÍSTICAS	25
3.3.1 - VALORES DOS <i>PIXELS</i>	25
3.3.2 - HISTOGRAMAS DE CINZA	26
3.3.3 - PCA (<i>PRINCIPAL COMPONENT ANALYSIS</i>).....	27
3.3.4 - LBP (<i>LOCAL BINARY PATTERNS</i>).....	28
4 - METODOLOGIA	30
4.1 - ORGANIZAÇÃO DOS DADOS.....	30
4.2 EXTRAÇÃO DE CARACTERÍSTICAS E CLASSIFICADORES UTILIZADOS	31
4.3 ANÁLISE ESTATÍSTICA.....	33
5 - RESULTADOS.....	35
5.1 - CARACTERÍSTICAS EXTRAÍDAS GLOBALMENTE.....	35
5.2 - CARACTERÍSTICAS EXTRAÍDAS LOCALMENTE.....	36
5.3 - ANÁLISE DO RESULTADO UTILIZANDO HISTOGRAMA DE CINZA	38
5.4 - ANÁLISE DOS RESULTADOS UTILIZANDO LBP	39
5.5 - ANÁLISE DOS RESULTADOS EM REGIÕES COM 7X7 PIXELS	39
5.6- ANÁLISE DOS RESULTADOS EM REGIÕES COM 3X3 PIXELS	40
5.7 - AVALIAÇÃO DE TODOS OS RESULTADOS USANDO ANÁLISE DE VARIÂNCIA (ANOVA).....	41
6 - CONCLUSÃO	46
REFERÊNCIAS BIBLIOGRÁFICAS	48

1 - INTRODUÇÃO

Nos dias atuais, as tecnologias são renovadas cada vez mais rapidamente e vão sendo inseridas na sociedade sem um tempo adequado entre elas. Em poucos meses é notório o avanço das tecnologias, seja de um celular, tablet, câmera fotográfica, computador, ou arquitetura de transferência de dados, etc. Sendo assim, alguns desafios e dificuldades ficam ultrapassados e novos vão surgindo na proporção em que essas evoluções vão aparecendo.

Este trabalho de pesquisa aborda o campo da visão computacional. A visão computacional engloba diversas subáreas, entre elas o processamento de imagem digital, classificadores de dados, reconhecimento de objetos, restauração de dados em imagens, entre outros. Mais detalhes na Seção 3.1.

Uma das subáreas da visão computacional é a da detecção de faces que ao contrário do que muitos pensam, não acontece naturalmente por parte da máquina, sendo necessário, antes de tudo, “ensiná-la” o que é cada objeto. Esse processo é denominado de Aprendizado de Máquina, discutido na Seção 3.2. Com a popularização de redes sociais, é facilmente visto nos álbuns de fotografias dos usuários o próprio sistema fazendo marcações automáticas das faces encontradas nas fotografias ou até mesmo câmeras digitais que possuem disparos automáticos da captura da fotografia de acordo com a expressão facial do indivíduo, ou seja, se as pessoas alvo da fotografia sorrirem, a máquina captura automaticamente a fotografia.

Tendo em vista que estas tecnologias avançadas estão cada vez mais próximas do nosso cotidiano, podemos pensar que esse tipo de abordagem é de fácil implementação por parte dos desenvolvedores e que a máquina facilmente fará seu trabalho. Antes que qualquer processamento possa ser realizado em imagens de faces, tais imagens devem ser detectadas.

A detecção de faces em imagens digitais tem se tornado cada vez mais abrangente e com diversas aplicações na sociedade. É possível verificar facilmente, nos dias de hoje, câmeras de segurança em estabelecimentos privados como restaurantes, hotéis, bancos, casas lotéricas etc. ou ambientes públicos como ruas, avenidas, semáforos etc. Essa prática tem por objetivo específico o monitoramento do ambiente, funcionários, transeuntes e qualquer eventualidade que ocorra. Se o objetivo do estabelecimento é gravar imagens de segurança para uma eventual análise das imagens gravadas a fim de buscar alguma informação ou

resolver um determinado problema que possa ter ocorrido, seria interessante, aliado a essa câmera de monitoramento de segurança, um software acoplado que fizesse reconhecimento das faces das pessoas no local (Pereira, 2008).

Seria interessante em um restaurante uma câmera com detecção de face capaz de fazer um reconhecimento das expressões das pessoas com o intuito de saber o nível de satisfação dos clientes presentes no estabelecimento. O mesmo aplicativo poderia fazer, por exemplo, um reconhecimento do cliente, para saber se ele frequenta sempre o ambiente e verificar em seu histórico de pedidos o que o cliente gosta ou ainda, dos últimos pedidos do cliente, quais ele se sentiu mais ou menos satisfeito. A mesma abordagem poderia ser feita em alguma loja de venda de produtos com uma câmera posicionada de frente para o cliente em uma vitrine com exposições de produtos e promoções. O papel desta câmera seria o de reconhecer o nível de satisfação, de acordo com as expressões faciais, das pessoas que param para analisar os anúncios.

Muitas aplicações podem ser feitas com o princípio da detecção de faces, mas, para isso acontecer, faz-se necessário a investigação de alguns fatores que afetam a detecção ao longo dos anos, tais como:

- **Oclusão** - Situação em que a face está coberta por algum objeto ou a pessoa está com algum acessório como, echarpes, boné, óculos de sol ou até mesmo um homem que possua barba. Esses fatores acabam impedindo ou dificultando a visualização limpa e clara da face.
- **Iluminação** - Pessoas que estão posicionadas em um ângulo de visão da câmera em que algum ponto de claridade provoca um sombreamento ou uma claridade excessiva na face.
- **Orientação** - Se dá quando o indivíduo está com a face posicionada para alguma direção que não seja frontal para a câmera de captura. Se o indivíduo está olhando para o chão, para os lados, de perfil, semiperfil e etc.
- **Qualidade da imagem** - Problema em que a câmera de captura não tem alto poder de resolução, dificultando, assim, a detecção por parte do software em função da baixa precisão que o software apresenta. Câmeras mal posicionadas, que dificultam a precisão do software na detecção das faces por conta da distância do objeto em questão.

Os eventos citados acima são os mais investigados na área da detecção de faces já que, sempre que qualquer aplicação for implementada, todos os problemas descritos deverão ser analisados antes do desenvolvimento do software de detecção de faces, que será acoplado à câmera que irá capturar as imagens.

Para ser possível a detecção de faces, existem alguns passos que devem ser seguidos:

1. **Percorrer a imagem** - O algoritmo será aplicado na imagem delimitando a áreas de busca.
2. **Extrair características** - Deve ser usado algum ou vários métodos de extração de características, a fim de capturar apenas as informações contidas na imagem.
3. **Classificação** - Os classificadores recebem as características extraídas (passo anterior) e classificam de acordo com o tipo de treinamento que lhe é passado (aprendizado supervisionado).
4. **Marcação** - Finalmente, com a região classificada, existe o processo de marcação. Se for encontrada alguma face por parte do classificador, é feita uma marcação (normalmente usa-se um quadrado de cor sólida) na imagem indicando qual lugar o classificador reconheceu como face.

Conhecendo os problemas enfrentados e quais processos são necessários para a realização da detecção de faces, faz-se necessário um estudo mais aprofundado e específico com o objetivo de analisar ou sugerir algum método que seja capaz de contornar algumas dessas situações.

Dos problemas expostos, o objetivo desta pesquisa tem foco exclusivo em analisar quais métodos de extração de características em imagens com faces que possuem vários tipos de iluminação e, conseqüentemente, sombreamento, têm o melhor resultado positivo de classificação combinado com dois tipos de classificadores. Foram utilizadas imagens do banco de dados *The Yale Face Database B¹*, que possui como característica imagens com sombreamento e iluminação em vários ângulos das faces.

Todos os testes foram analisados estatisticamente utilizando ANOVA, cujo objetivo é verificar se existe diferença significativa entre os métodos. Os cálculos da ANOVA foram feitos utilizando uma planilha eletrônica que oferecia essa análise dos dados automaticamente.

¹ <http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html>

Fica claro que o objetivo desta pesquisa não foi propor nenhum método novo de extração de características e classificador, mas avaliar o que já existe de melhor aplicando-os a um tipo de problema específico comumente encontrado na detecção de face.

2 - REVISÃO BIBLIOGRÁFICA

Nesta seção serão descritos alguns trabalhos que serviram como base para o avanço da detecção de faces e apresentará alguns dos recentes estudos realizados na área.

Para um melhor entendimento serão descritos dois dos métodos de classificação de característica que são bastante utilizados no reconhecimento de padrões. São eles: Haar e Adaboost.

O método Haar, também conhecido como *HaarCascade*, foi proposto por Viola e Jones (2001a) adaptando a abordagem criada inicialmente por Papageorgiou et. al. (1998). Tem por característica delimitar a área de busca em pequenos retângulos onde as características são extraídas e, posteriormente, é comparada com a imagem alvo. Podem atuar em cascata e com vários conjuntos de classificadores com fraco desempenho de classificação, porém sua organização final acaba sendo robusta.

O método AdaBoost (Adaptive Boosting), é um algoritmo de aprendizado de máquina, criado por Freund e Schapire (1997). Sua característica permite o acoplamento com outros tipos de algoritmos de aprendizagem, melhorando seu desempenho. Além disso, os classificadores se adaptam de acordo com o resultado dos classificadores anteriores. Seu conjunto pode ser composto por diversos classificadores de fraco desempenho, ainda assim ele consegue criar um modelo final bem representado.

Os dois métodos citados acima possuem características muito peculiares e são bastante utilizados, principalmente pela necessidade do reconhecimento em tempo real de algumas aplicações e sua capacidade de acoplamento com outros classificadores, possibilitando o estudo de novos métodos e combinações do que existe de melhor. Alguns trabalhos citados a seguir, demonstram como essas combinações entre classificadores podem ser aplicadas.

Gorbenko e Popov (2012) afirmam que utilizar classificadores do tipo haar cascade na detecção de faces em tempo real, como abordam Viola e Jones (2001a) idem (2001b), é bastante frequente, porém, a esse uso deve-se acoplar com algum outro tipo de detector específico para a finalidade proposta, como auxílio da detecção. Sendo assim, para uma

detecção de faces eficaz, se faz necessário um algoritmo complexo que combine o que cada método distinto tem de melhor.

Uma abordagem proposta por Pereira et. al. (2011), comprova que as combinações de vários classificadores utilizados para treinamento obtiveram resultados mais satisfatórios do que os classificadores treinados e testados isoladamente. As características foram extraídas das imagens de modo global e local assim como foi feito neste trabalho (Seção 4.2). No modo de extração local, foram realizadas e avaliadas várias métricas a fim de verificar quais regiões seriam candidatas à extração de características por possuir dados mais significativos na imagem. Das métricas avaliadas foram analisados tamanhos e formas das regiões e as relações de valores entre elas, teste com validação cruzada, ajuste do custo e gama no kernel-SVM. Das imagens utilizadas, continham face com poses frontais (incluindo semiperfis) e perfis, as taxas de acerto foram de 91,7% e 99,59%, para verdadeiros positivos e verdadeiros negativos, respectivamente.

Ribeiro et. al. (2012) apresentam um conjunto de detectores em cascatas baseados no IPD (Inner Product Detector) capaz de fazer a detecção de pontos fiduciais sobre a face em tempo real. O sistema implementado é robusto a variações de iluminação, translações e pequenas rotações. Foram utilizados 13 pontos sobre a face de forma manual para treinamento. O treinamento utilizando o classificador IPD permite definir parâmetros da cascata de detectores de maneira objetiva. O sistema foi implantado no formato modular, ou seja, era possível a utilização de outros detectores específicos desde que esses oferecessem uma variedade de candidatos a pontos fiduciais para o detector seguinte. Cada detector tem por objetivo analisar os pontos propostos como candidatos dos classificadores anteriores e eliminar os de menor ocorrência. Foram impostos dois critérios de satisfação de parada na cascata de classificadores: a quantidade máxima de cascata de acordo com a taxa de acerto e o outro considera o custo computacional de adicionar uma nova cascata. A métrica de avaliação dos pontos candidatos foi feita pela avaliação dos rótulos manuais e os feitos automaticamente. Por ser rápido e de baixa complexidade, o sistema tornou-se viável para aplicação em tempo real, condições de iluminação distintas e pequenas rotações da face no plano.

Uma abordagem semelhante foi proposta por Veloso et. al. (2007) onde a partir da detecção de faces é feito o reconhecimento da expressão facial detectada. São realizadas

marcações manuais dos pontos na face com a finalidade de mapear características relevantes de acordo com cada expressão facial. Ao contrário da proposta feita por Ribeiro et. al. (2012), o estudo feito por Veloso et. al. (2007) foi mais adiante e procurou classificar qual expressão facial o indivíduo está fazendo, como alegria, tristeza, surpresa, medo, raiva e nojo.

Uma implementação feita por Quintans et. al. (2010) usando *HaarCascade* através do OpenCV concluiu que a variação de luz e a baixa qualidade da webcam interferiram totalmente nos seus experimentos, embora a detecção parcial (nariz, olhos e boca) tenha sido realizada, houveram momentos em que nem todas as partes foram detectadas simultaneamente em função do nível de exposição da luz na face, interferindo no resultado final.

Uma abordagem proposta por Baioco e Salomão (2010) tem como objetivo usar uma rede neural como classificador na detecção de faces em multidões. É possível verificar que, com ajustes das imagens, como a equalização do histograma, o tipo de extração de característica, o tipo de classificador, o tipo e a qualidade da base de imagens e até mesmo o tipo de compressão da imagem, há interferência no resultado obtido por parte do classificador, ficando claro que simples alterações na imagem podem fazer total diferença na detecção de faces.

Os estudos feitos por Quintans et. al. (2010) e Baioco e Salomão (2010) são um típico exemplo que reforça a importância desta pesquisa de analisar a influência de fatores de iluminação e sombreamento na imagem que possam interferir em alguma aplicação específica. Enquanto Quintans et. al. (2010) encontraram sérias dificuldades com iluminação, por exemplo, Baioco e Salomão (2010) contornaram o problema com apenas alguns ajustes na imagem, aumentando a taxa de acerto em 20% aplicando apenas um filtro de processamento de imagem. Episódios como esse são muito comuns, com isso, surge a necessidade de análise e mapeamento de alguns problemas frequentes e desenvolvimento ou sugestão de como melhorar o desempenho na detecção de faces para um determinado fim.

Carvalho e Tavares (2008) propuseram uma abordagem para detecção de faces baseada no método de detecção do tom da pele e uma segunda baseada em um modelo protótipo para detecção dos olhos como validação da primeira. No primeiro caso é feita uma segmentação da imagem considerando os valores dos seus pixels e com análise probabilística é feita uma comparação das imagens de tons de pele com as imagens testadas. Como esse

método não atribui confiança suficiente para a detecção de faces é realizado um complemento utilizando um detector de olhos que será aplicado na região em que o tom de pele foi considerado positivo na imagem a fim de procurar por olhos naquela região validando ou não a detecção anterior. Para compor o método estatístico foram usados 16 tons de pele diferentes. No segundo processo, em cada imagem testada foi realizada uma busca na imagem segmentada a procura de lugares 'vazios' ou sem informação (mesma cor), no caso, a busca consistia em achar três (representaria dois olhos e uma boca) regiões vazias com um limiar de 500 pixels (proporcional à imagem testada nos experimentos) vizinhos que contivessem a mesma cor. A partir daí, era feita a aplicação da detecção de olhos como reforço do primeiro procedimento. Os resultados desta combinação se mostraram satisfatórios com elevada taxa de confiança. Todos os testes foram feitos utilizando matlab para uma única imagem por vez ficando, assim, uma necessidade de testes mais robustos.

Assim como novas técnicas de classificação e acoplamentos entre si vão surgindo, o mesmo acontece com alguns métodos de extração de características que possibilitam essa integração com outros métodos e variação da sua forma original. É o caso do extrator de característica proposto por Pereira et. al. (2010) que incrementa o método do LBP proposto por Ojala et. al. (1996). Esse método foi chamado de INTLBP (*Integral Local Binary Pattern*) e utiliza-se da técnica da abordagem de imagem integral para representar histogramas de LBP. Essa técnica aumenta a velocidade de processamento na detecção de faces e aumenta a taxa de acerto comparado com a abordagem do LBP Invariante a Rotação (Ojala et. al., 2002) como afirma Pereira et. al. (2010).

Prado e Gonzaga (2012) abordaram um método inovador de detecção de faces baseado em proporções geométricas denominado de proporção áurea. Essas proporções podem ser representadas, por exemplo, como segmentos ou retângulos. A proporção Áurea ou Divina é amplamente utilizada na área médica para fins estéticos. Nesse método é feito um pré-processamento em busca de bordas na imagem e com a imagem binarizada é aplicado um filtro para diminuição do ruído, diminuindo partes desconexas com número de pixel superior a 5 aglomerados distintamente de outro grupo de pixels. Em seguida, é feita a extração de características utilizando o método proposto que é aplicado ao classificador utilizando lógica Fuzzy implementado em Matlab. Todos os testes encontraram dificuldades com iluminação, brilho, oclusão, tamanho da face e da imagem entre outros. Esse método é inovador por buscar faces uniformes e proporcionais diretamente na imagem, sua taxa de acerto foi de

96%, valor bastante próximo de resultados usando técnicas conhecidas. O lado ruim é que se uma face em teste estiver fora dos padrões simétricos, ou seja, seu rosto tenha alguma má formação ou desproporcional, teoricamente sua probabilidade de acerto é quase nula. Os testes realizados incluíram 100 imagens, com isso, é necessário uma melhor investigação e aumento da quantidade e variações das faces.

Outra abordagem interessante é a feita por Silva et. al. (2012) que propõe uma metodologia na detecção de faces frontais humanas submetidas a diversas condições de iluminação, textura, tamanho, etc. baseada na textura da boca do indivíduo da imagem, tentando auxiliar na diminuição dos erros de detecção e reconhecimento da face. Como método de extração de característica foi adotado a matriz de coocorrência proposta por Haralick e Shanmugam (1973), que armazena em uma matriz a transição de níveis de cinza entre os pixels que posteriormente serão usados para fins de análises probabilísticas de cada ocorrência. Os parâmetros estatísticos desse método são considerados de segunda ordem por não serem derivados diretamente de dados das imagens como afirmam Pham e Alcock (1998). Devido a matriz criada ter ordem de tamanho proporcional ao maior nível de cinza da imagem, seu custo de processamento torna-se alto, embora a técnica tenha demonstrado bastante eficiência para o fim proposto que é fazer detecção de boca em imagens antigas escaneadas. Foram analisadas, inicialmente, dos dados das imagens de boca as seguintes características: contraste, entropia, segundo momento angular, correlação, variância e homogeneidade, obtidas através da matriz de coocorrência. O classificador utilizado foi o SVM em 10432 imagens e obteve uma taxa de acerto de 84% para imagens de face.

Apesar das inúmeras formas possíveis para a realização da detecção de faces citadas acima, nenhuma trata especificamente do problema da iluminação e oclusão ou propõe um método adequado para este problema ou um método com alto nível de confiança semelhante ao que existe nos métodos aplicados em imagens frontais em condições perfeitas (boa iluminação e sem oclusão). Alguns dos trabalhos relacionados acima ficariam com suas taxas de acerto comprometidas, caso as imagens de face usadas nos experimentos, contivessem iluminação inadequada ou oclusão, são os casos de Ribeiro et. al. (2012), Veloso et. al. (2007), Carvalho e Tavares (2008) e Prado e Gonzaga (2012) que necessitam de toda a área da face para aplicações de seus métodos propostos. Com isso, surge a necessidade de saber se os métodos de extração de características existentes combinados com algum classificador supre esse problema de forma satisfatória, sem a necessidade de criar um novo método.

3 - FUNDAMENTOS TEÓRICOS

3.1 - VISÃO COMPUTACIONAL

Jain et al. (1995) descrevem o campo da visão computacional como a área da computação que tem por objetivo simular a percepção do olho humano em máquinas. Seus algoritmos estão intimamente ligados a áreas como a física e a matemática. Dentre as subáreas abordadas diretamente ou indiretamente estão o processamento digital de imagem, a detecção e reconhecimento de padrões e objetos, o aprendizado de máquina, a inteligência artificial, a restauração de imagens, entre outras. A visão computacional surgiu de problemas encontrados em outras áreas, e não possui uma formulação exata de seus problemas, assim como as soluções deles. Com isso, os mapeamentos dos problemas classificados como sendo de visão computacional, acabam ficando restritos a situações muito particulares. Essa área, comparada a outras, é relativamente nova, tendo em vista que seu surgimento se deu após o desenvolvimento do computador e sua popularização, com isso ainda existem muitos paradigmas, dificuldades e problemas que necessitam ser investigados e contornados. Mais detalhes em Jain et al. (1995).

3.2 - APRENDIZAGEM DE MÁQUINA

De acordo com Monard e Baranauskas (2003), o aprendizado de máquina é um subcampo da inteligência artificial que tenta, por meio de algoritmos específicos e sofisticados, tentar reproduzir percepções humanas nas máquinas, a fim de que elas possam tomar decisões por conta própria através de modelos representativos em busca de resolver algum problema ou executar alguma tarefa. Ele está diretamente ligado à refinação e análise dos dados, métodos estatísticos e do conceito indutivo, ou seja, faz inferências aos dados do problema partindo de conclusões que ele já conhece. Com isso, surgem alguns problemas típicos da indução. Por exemplo, é possível afirmar que o ferro é condutor de eletricidade e é um metal, que o cobre é um condutor de eletricidade e é um metal, que o zinco é um condutor de eletricidade e é um metal, para uma máquina que contivesse esses dados de entrada, ao analisar que a água conduz eletricidade, então a máquina aprendiz poderia classificar que a água também é um metal. Ainda assim, existem resultados bastante satisfatórios nas áreas de sua aplicação e os algoritmos tem evoluído constantemente tratando dos erros e exceções com

bastante robustez. Para melhor entendimento no assunto é recomendado a leitura de Monard e Baranauskas (2003).

Partindo deste princípio, foram utilizados classificadores nesta pesquisa que se utilizam de modelos estatísticos e de análises dos dados de entrada a fim de representá-los em sua aprendizagem supervisionada, de modo que sejam capazes de fazer as classificações de objetos distintos não contidos em seu modelo de aprendizado.

Os classificadores utilizados nos treinamentos e testes com as características extraídas dos métodos são listados logo abaixo:

- Máquinas de Vetores de Suporte (*Support Vector Machine – SVM*);
- Redes Neurais Artificiais do tipo *Perceptrons* de Múltiplas Camadas (*Multi-layer Perceptron – MLP*).

3.2.1 SUPPORT VECTOR MACHINES (SVM)

As máquinas de vetores de suporte são excelentes ferramentas para o reconhecimento de padrões e utilizam o método de aprendizado supervisionado. A SVM trabalha com modelos de entrada, ou seja, o utilizador da SVM define o que ela tem que saber, quais dados serão passados para ela e, com isso, espera-se que a SVM seja capaz de tomar decisões em dados desconhecidos a partir do seu modelo inicial, gerando uma saída (resposta) com o resultado de sua análise. Para comparação dos dados que terá de tratar, a SVM utiliza-se da análise estatística detalhada dos dados. Seu algoritmo foi baseado no modelo desenvolvido por Vapnik (1995), que possui um método de classificação bastante confiável e bem sucedido. Com isso, seu uso se tornou bastante aceitável no meio acadêmico, científico e de pesquisa. Segundo Lorena e Carvalho (2003) o ponto forte da SVM é sua capacidade de generalização dos dados e sua forma de separação das classes dos dados treinados e testados. As SVMs têm o poder de generalização nos planos 2D e 3D se necessário, o que dá forte consistência e robustez a essa ferramenta podendo, assim, receber uma carga maior de dados com pouca variância entre eles, que seria capaz de representar a divisão das classes. A desvantagem é por conta da sua velocidade em todas as etapas de sua utilização. Quanto maior o número de dados, mais demorado é o processo. Também, se a variação dos dados entre as classes forem pequenas, será necessário um maior número de vetores de suporte para separação e definição

das classes, o que torna o processo lento. Para mais detalhamento sobre SVM recomenda-se a leitura de Tahim (2010) e Lorena e Carvalho (2003).

3.2.2 REDES NEURAIAS ARTIFICIAIS (RNA)

As Redes Neurais Artificiais (RNA) são fundamentadas em modelos computacionais baseados em ligações semelhantes ao modelo cerebral, ou seja, existe uma estrutura com capacidade de realizar operações as quais estão ligadas entre si a fim de compartilharem informação/dados com algum objetivo. A RNA tem a característica de aprendizado através do fornecimento de dados, ou seja, assim como as SVMs, as RNAs também têm em sua forma o aprendizado supervisionado, como apontam Portugal e Fernandes (1996). Chama-se aprendizado supervisionado pelo fato do usuário quem decide o que a máquina deve aprender e quais objetos pertencem a qual classes. Os benefícios do uso das redes neurais artificiais (RNA) se dão pelo fato da sua capacidade de fazer diversas operações em tempo reduzido; criando sua própria representação da informação de entrada; tem tolerância a falhas, isto é, se uma parte da RNA está danificada, ela tem a capacidade de conseguir dar continuidade nas operações sem comprometer a rede por inteiro, como afirmam Dias e Radonsky (2003). Isso se dá pelo fato de armazenar informações redundantes, além disso, a RNA tem a capacidade de modelar qualquer função determinada, isso se dá pelo fato da quantidade de informação que a RNA pode obter em sua estrutura. O problema da RNA está na sua sensibilidade à variação dos dados (Haykin, 1998). Se o treino e validação forem feitos com um tipo de imagem, qualquer alteração nos dados de testes pode ser comprometedora para a rede neural. Com isso, existe a recomendação de vários treinos com validação em busca do melhor modelo representacional para aquele conjunto de dados.

Na Figura 1 é apresentado um modelo representacional de uma RNA, onde é possível verificar a existência de três camadas: a camada de entrada, as camadas intermediárias e a camada de saída. Cada elipse representa um nó e todos os nós de uma camada estão ligados entre si e com os nós das próximas camadas, todos contêm valores atribuídos aleatoriamente, assim como cada ligação entre os nós. Para a finalidade desta pesquisa, a camada de entrada recebe os valores das características extraídas dos métodos, ou seja, se estivermos utilizando como método de extração de característica os valores dos pixels, em uma imagem de 20x20 serão extraídas 400 características, logo, a camada de entrada terá

400 nós de entrada, isso significa dizer que para cada característica deve-se existir um nó de entrada.

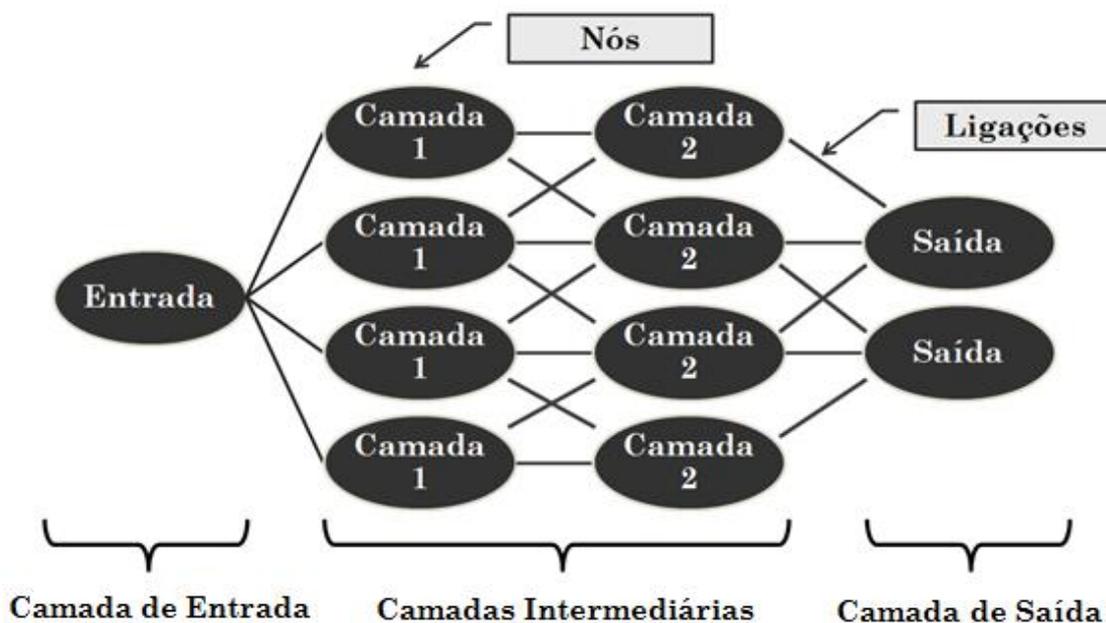


Figura 1 - Exemplo de uma rede neural artificial

(Fonte: acervo do autor)

Na camada intermediária, podem existir diversas subcamadas e a quantidade de nós é definida pelo criador da rede. Essa camada também é chamada de camada escondida. Essas subcamadas têm por função fazer cálculos entre os dados de entrada com valores aleatórios atribuídos a cada nó e ligação. Se os cálculos de soma e multiplicação dos dados forem maiores que o limiar atribuído pelo criador da rede neural, então aquele nó propaga em sua saída um sinal que o cálculo excedeu o limite estabelecido. As saídas da rede neural são avaliadas em relação à saída desejada. A avaliação ocorre por meio de medidas como o RMSE (Root Mean Square Error) que consiste no seguinte cálculo: cada saída da rede é subtraída do valor de saída desejado; em seguida, o resultado da subtração é elevado ao quadrado; calcula-se à média dos quadrados e extrai-se a raiz quadrada da média. Essas diferenças são propagadas de volta para a rede para ajustar os pesos. Foi utilizada nos experimentos desta pesquisa a rede neural multilayer perceptron com algoritmo de treinamento backpropagation² e função de ativação³ sigmoidal⁴.

² http://www.lncc.br/~labinfo/tutorialRN/frm4_backpropagation.htm

³ http://en.wikipedia.org/wiki/Activation_function

⁴ http://en.wikipedia.org/wiki/Sigmoid_function

Os valores iniciais de cada nó e ligações são aleatórios e normalmente pequenos. A medida que a rede vai recalculando, os valores vão sendo modificados e consequentemente aumentados. Na proporção que os cálculos de alguns nós vão sendo mais utilizados que outros, estes se tornam mais importantes para a decisão enquanto os nós menos utilizados vão enfraquecendo e diminuindo seus valores. A camada de saída representa a quantidade de classes definidas na entrada, no caso desta pesquisa estamos trabalhando com imagens de faces e não faces, logo, existem duas classes uma para face e outra para não face. Consequentemente, a camada de saída possuirá dois nós. Para estudos mais aprofundados em redes neurais artificiais recomenda-se a leitura de Haykin (1998).

3.3 - PROCESSAMENTO DE IMAGENS: EXTRAÇÃO DE CARACTERÍSTICAS

No campo do processamento de imagens e reconhecimento de padrões, após a extração de características é aplicado um determinado algoritmo cujo objetivo é reduzir a quantidade dos dados de entrada, ou seja, a ideia é extrair algumas características de uma imagem, ao invés de utilizar todas as informações presente nela. Pode-se utilizar uma quantidade reduzida das características, eliminando, assim, informações redundantes, diminuindo o tempo e a carga de processamento por parte do computador. (Nixon e Aquado, 2012).

Os métodos de extrações de características utilizados para treinamentos e testes nos classificadores estudados neste trabalho foram:

- Valores dos *Pixels*;
- Histogramas de níveis de cinza;
- Análise de Componentes Principais (*Principal Component Analysis* – PCA);
- Padrões Binários Locais (*Local Binary Patterns* – LBP);

3.3.1 - VALORES DOS PIXELS

Este método de extração de características trabalha com o valor de todos os pixels de uma região da imagem ou da imagem completa que será processada, sendo assim, não existe redução da dimensionalidade da imagem como também não existe eliminação das características redundantes em regiões próximas. Se uma imagem possui dimensões de 21x21 a quantidade de pixels dessa imagem será de um total de 441 pixels que consequentemente

será o total de características obtidas para o processamento do classificador. Nota-se que quanto maior a dimensão da imagem (x,y) maior será a quantidade de informação a ser tratada, o que implica numa maior carga de dados, aumentando o tempo de processamento. Dos quatro métodos adotados nesta pesquisa este é o único que não tem uma redução da dimensionalidade e não elimina as informações redundantes na imagem.

3.3.2 - HISTOGRAMAS DE CINZA

Um histograma é uma representação da frequência de ocorrências de valores. Para uma imagem, um histograma pode ser usado para contar a ocorrência dos valores das intensidades dos pixels. Cada número que corresponde à quantidade de ocorrências pode ser chamado de bin. Um histograma também pode ser usado para verificar a ocorrência de valores em faixas pré-determinadas. Se dividirmos uma imagem em escala de cinza que possui 256 representações (zero a 255) por 16 bins, os resultados serão 16 blocos com 16 níveis de cinza para cada bloco. Por exemplo, para calcular histogramas para uma imagem em tons de cinza que utiliza 8 bits para representar os 256 níveis de intensidade diferentes podemos usar diversas quantidades de bins. Se usarmos, por exemplo, 8 bins, cada bin conterà a quantidade de vezes que os valores em faixas com 32 níveis ocorreram. No caso de 8 bins, as faixas seriam: 0 a 31, 32 a 63, 64 a 95, etc. A Figura 2 (a) exibe um exemplo de uma imagem com 4x5 pixels e seus respectivos valores. A Figura 2 (b) representa a distribuição dos bins para um histograma com 4 faixas de valores.

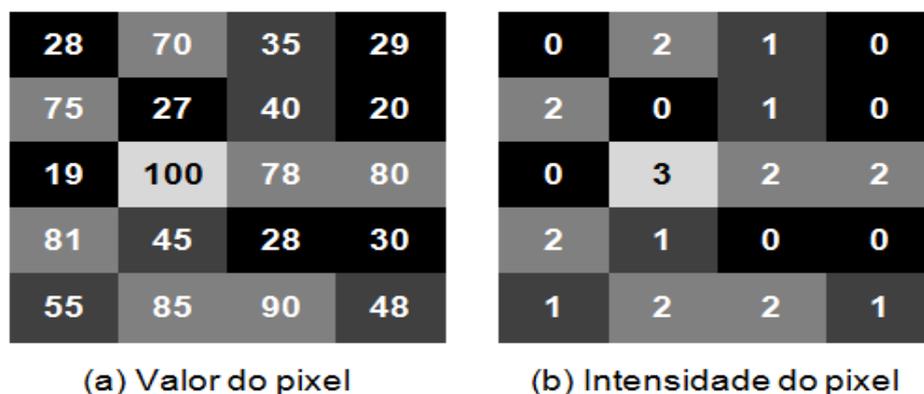


Figura 2 - Mapeamento entre faixas de valores (bins) e posição dos pixels na imagem
(Fonte: Acervo do autor)

Na Figura 2 (b) é apresentado o quociente do cálculo da divisão do valor do pixel pela quantidade dos níveis (32 nesse caso) que cada bin representa utilizando 8 bins, ou seja, ao dividir 28 por 32 o quociente da divisão é menor que 1, logo, o bin zero recebe um

incremento de valor 1 em seu contador. Após o cálculo da intensidade de cada pixel, é feita a quantização desses valores por faixas de níveis, como mostrado no gráfico 1 abaixo. É visto que o bin zero possui sete pixels com intensidade zero, por outro lado, a maior intensidade (de cor mais clara), possui apenas uma representação no bin 3.

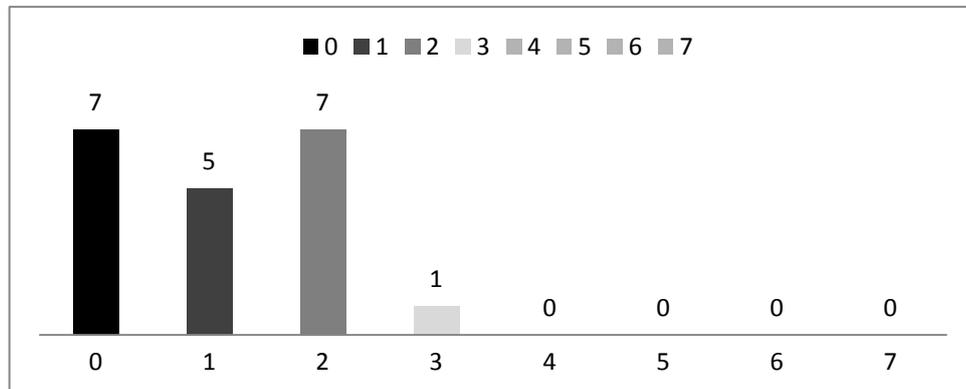


Gráfico 1 - Representação gráfica de um histograma de níveis de cinza.

Como é visto no Gráfico 1, os valores dos bins 4, 5, 6 e 7 possuem representações insignificantes, pois seus conteúdos estão vazios.

3.3.3 - PCA (*PRINCIPAL COMPONENT ANALYSIS*)

A Análise de Componentes Principais (PCA) foi criada por Pearson (1901) e consiste em utilizar combinações lineares de forma a reduzir significativamente a quantidade representacional dos dados extraídos de alguma fonte. Sendo assim, a quantidade de características utilizada poderá ser diminuída de acordo com a escolha do usuário. As principais etapas para a realização do PCA, segundo Vasconcelos (n.d.) são: o cálculo das médias dos dados obtidos e, posteriormente, a subtração de cada valor dos dados da média calculada. Feitas essas operações, será calculada a matriz de covariância, utilizando essas subtrações. A matriz de covariância tem por conceito generalizar um conjunto de variáveis aleatórias e verificar a média de acoplamento entre elas.

O próximo passo é calcular os autovetores de comprimento 1 (não nulos) e autovalores. Os autovalores são os valores representativos de uma matriz que visam resumir ou simplificar as propriedades daquela matriz. Para os autovalores serem aceitos é necessária a subtração do valor encontrado pelos valores de cada elemento da diagonal da matriz principal, de forma que essa se torne singular. Os autovetores são os vetores encontrados de

operações lineares a fim de garantir a preservação da direção do vetor. Eles estão ligados aos autovalores – cada operador linear está ligado a um conjunto específico de autovetor.

Por último, a matriz é organizada de forma que os valores mais significativos são dispostos de forma decrescente, ficando assim uma lista dos mais importantes até os menos significativos. Com isso, é possível obter um número reduzido de padrões de característica da imagem, melhorando o desempenho no processamento e eliminando as informações redundantes.

3.3.4 - LBP (*LOCAL BINARY PATTERNS*)

O LBP é um algoritmo de extração de características bastante robusto quando aplicado em detecção de faces, tendo em vista sua facilidade de manipulação e integração com outras técnicas. O LBP é bastante utilizado, tendo em vista sua invariância à iluminação. O LBP original proposto por Ojala et al. (1996) é baseado em valores dos vizinhos do pixel central. Uma máscara com 3x3 é utilizada com um valor pré-definido do raio de ação dos vizinhos que varre a imagem a fim de atribuir valores de zero e um a cada valor do pixel. Se o valor do pixel do vizinho for menor que o valor do pixel central é atribuído o valor 1 àquela posição do pixel, caso contrário, será atribuído zero. A disposição dos zeros e uns garante nesse método a contagem de transições entre eles, fazendo dessa transição um padrão que é adotado pelo classificador.

A Figura 3 exemplifica uma máscara de 3x3 realizando o método de extração de característica LBP em uma imagem. Na Figura 3 (a) são mostrados os valores originais dos pixels e o pixel central escolhido para comparação com os vizinhos. Na Figura 3 (b) é exibido o conteúdo após a aplicação do LBP, em que quando o valor do vizinho do pixel central foi menor que o pixel central, este recebeu valor 1, caso contrário, recebeu valor zero. Contando os números binários no sentido horário a partir da posição 1, se obtém a seguinte sequência: 11101100, que transformando de binário para decimal ficará: 236. Este será o novo valor na matriz de valores, como mostra a Figura 3 (c). A grade de análise passará para o valor seguinte até que todos os números sejam atualizados.

0	4	3
40	5	15
5	2	1

(a) Representação dos valores dos pixels

1	1	1
0		0
0	1	1

(b) Valor em binário

1	2	3
9	<u>236</u>	6
5	3	1

(c) Atualização dos Valores

Figura 3 - Exemplo do método LBP
(Fonte: Acervo do autor)

A abordagem do LBP utilizada nesta pesquisa inclui a aplicação do método de histograma de cinza (descrito na subseção anterior) na matriz de números atualizada. Como há diversas variações para extração de características LBP, recomenda-se a leitura de Pereira et. al. (2010), Ojala et. al. (1996), Ojala et. al. (2002), Rodriguez (2006), Marcel et. al. (2007) para aprofundamento no tema.

4 - METODOLOGIA

Neste capítulo, será apresentada a metodologia necessária para a realização deste trabalho. Na Seção 4.1 serão apresentados como os dados foram obtidos e organizados para a realização dos experimentos. A Seção 4.2 lista os métodos de extração de característica nos modos globais e locais, assim como as bibliotecas utilizadas para a implementação dos classificadores. A Seção 4.3 descreverá como os experimentos foram validados estatisticamente.

4.1 - ORGANIZAÇÃO DOS DADOS

Como os principais fatores estudados nesta pesquisa foram os vários níveis de iluminação que podem influenciar no sucesso ou fracasso na detecção de faces, várias bases de dados foram analisadas a procura da que melhor se adequava aos nossos objetivos. A base escolhida foi a *The Yale Face Database B* composta por imagens de face com vários níveis de iluminação em vários pontos diferentes na face, conseqüentemente formando uma sombra no lado oposto do foco da iluminação.

Para o início das atividades, foi necessário à criação de dez pastas, cada uma contendo duas pastas, uma recebeu o nome de 'treino' e a outra de 'teste'. A ideia da divisão em dez grupos se dá pela necessidade de existir dez tipos de testes diferentes variando seu conteúdo de treino para validação estatística e comprovar a diferença entre os dados. O processo de seleção e separação das imagens foi feito cuidadosamente garantindo que nenhuma das imagens que estavam contidas na pasta "treino" estivessem repetidas na pasta "teste". Da mesma forma, nos dez grupos de pasta, o treino da pasta 1 foi elaborado de modo que nenhuma imagem que estivesse contida na pasta "treino" não estivesse em nenhuma outra pasta "treino" dos nove grupos restantes. Isso garante que tenho dez grupos distintos de imagens para treinamento.

O mesmo procedimento aconteceu para a separação das imagens de não face e foi utilizada como base de dados a disponibilizada por Naotoshi Seo em sua página sobre treinamento de cascatas de classificadores⁵. Em cada um dos dez grupos existentes foram criadas novas pastas dentro de cada pasta "treino" e "teste" que recebeu o nome de "não face". Como proporções de quantidade de imagens em cada pasta, foi adotada a seguinte medida: todas as pastas de treino receberam 1250 imagens de faces e 23mil imagens de não face. As

⁵ <http://tutorial-haartraining.googlecode.com/svn/trunk/data/negatives/>

pastas de testes ficaram com 3140 imagens de face e 120 mil imagens de não face. O grande número de imagens de não face se dá pelo fato da necessidade de tentar abranger os diferentes tipos de objetos, pedaços de faces, texturas e etc. para garantir a maior representação real possível de tudo que não é face. Todas as imagens foram redimensionadas para o tamanho 21x21 pixels garantindo a mesma quantidade de características extraídas das duas classes (face e não face).

A próxima seção apresentará os métodos de extração de características utilizados nesta pesquisa, assim como os classificadores utilizados.

4.2 EXTRAÇÃO DE CARACTERÍSTICAS E CLASSIFICADORES UTILIZADOS

Após a organização das imagens em conjuntos distintos, foi iniciado o processo de extração de característica das imagens para os testes propostos.

Todos os métodos aqui citados foram implementados usando linguagem de programação C/C++ utilizando a biblioteca OpenCV⁶ que continha o suporte necessário para o processamento dos conteúdos das imagens. Toda esta pesquisa foi feita utilizando o sistema operacional Linux.

Dos métodos utilizados como extração de características foram utilizados quatro tipos:

- Valores das intensidades dos pixels;
- Histogramas de imagens em escala de cinza;
- Análise de Componentes Principais (*Principal Componente Analysis – PCA*);
- Padrões Binários Locais (*Local Binary Patterns - LBP*) (Ojala et. al. (1996)).

Todos esses métodos foram descritos nas Seções 3.3.1, 3.3.2, 3.3.3 e 3.3.4. Cada método foi analisado e estudado isoladamente dos outros, ficando assim, um método aplicado e testado por vez. Os métodos foram aplicados em todos os dez grupos do conjunto, com isso cada método obteve dez resultados diferentes.

Para cada método de extração de característica foram designados para fins de categorização das imagens dois classificadores:

- Máquinas de Vetores de Suporte (*Support Vector Machine – SVM*);

⁶<http://opencv.org/>

- Redes Neurais Artificiais do tipo *Perceptrons* de Múltiplas Camadas (*Multi-layer Perceptron* – MLP).

Os classificadores utilizados foram descritos nas Seções 3.2.1 e 3.2.2 Para implementação dos classificadores foram utilizadas as seguintes bibliotecas:

- LibSvm⁷: *A library for support vector machines* (Uma biblioteca para máquinas de vetores de suporte);
- LibFann⁸: *Fast artificial neural network library* (Biblioteca de redes neurais artificiais rápida).

Todos os métodos de extração de características utilizaram os dois tipos de classificadores nos dez experimentos feitos (um para cada grupo do conjunto de dez pastas).

As características foram extraídas das imagens de duas formas:

- Local - A imagem é dividida em regiões contendo partes iguais, onde o método extrator de característica é aplicado em cada uma das regiões. Os dados extraídos são concatenados sequencialmente formando o padrão da imagem correspondente. A ordem de concatenação pode ser vista na Figura 4.
- Global - O método extrator de característica é aplicado na totalidade da imagem. Ficando, assim, sem a necessidade de concatenação, pois a própria imagem completa gerará seu padrão de uma única vez.

Dos métodos que usaram características globais: valores dos Pixels e PCA. O método de valores dos pixels foi adotado por utilizar todos os valores de todos os pixels da imagem, conseqüentemente já tem por sua característica a abordagem global. O PCA foi adotado nas características globais, pois sua forma de atuação e obtenção dos principais conteúdos da imagem é bastante positiva. Os demais métodos possuem em sua formação a característica de atuação local, ou seja, tem menor representatividade dos dados necessitando da divisão da imagem em regiões.

Na abordagem local todos os métodos de extração de características foram adotados, exceto os valores dos pixels (por usar sempre a totalidade da imagem). As regiões foram divididas em duas formas: imagens com regiões de 7x7 pixels totalizando 9 regiões em uma imagem de 21x21 pixels (Figura 4 (b)) e regiões com 3x3 pixels, totalizando 49 regiões em

⁷ <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

⁸ <http://leenissen.dk/fann/wp/>

imagens com 21x21 pixels, como é visto na Figura 4 (a). Com exceção da utilização do PCA que seu poder obtenção de dados é maior, as imagens foram convertidas para dimensões de 20x20 pixels e conseqüentemente as regiões foram distribuídas em tamanhos de 10x10 pixels, totalizando 4 regiões em uma imagem com 20x20 como pode ser visto na Figura 4 (c).

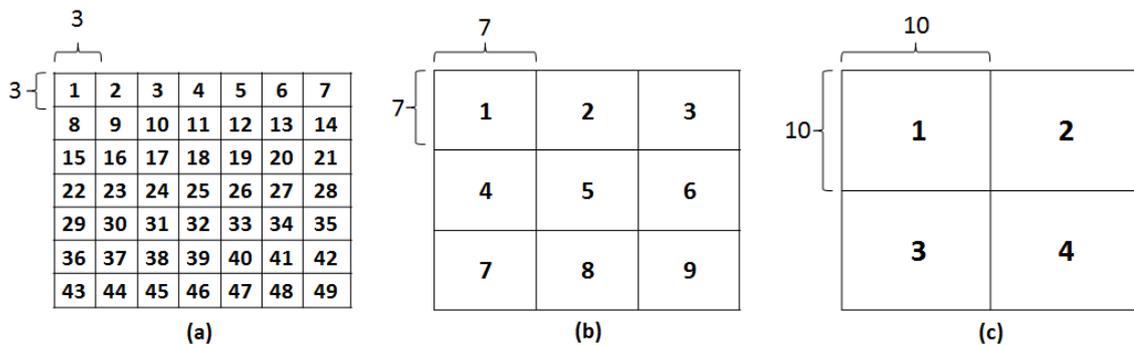


Figura 4 - Quantidade de regiões e ordem de concatenação

(a) - Representação de uma imagem que possui regiões de 3x3 pixels, totalizando 49 regiões; (b) - Representação de uma imagem que possui regiões de 7x7 pixels, totalizando 9 regiões; (c) - Representação de uma imagem utilizada pelo método PCA extraíndo as características locais que possuem regiões de 10x10 pixels, totalizando 4 regiões.

(Fonte: acervo do autor)

4.3 ANÁLISE ESTATÍSTICA

O método de divisão em dez grupos ocorreu pela necessidade da avaliação dos dados e conseqüentemente verificar se os mesmos diferem entre si. Para este propósito foi utilizado um método estatístico que analisa a variância dos dados e dependendo da diferença entre eles a pesquisa é validada ou não. Este método recebe o nome de Análise de Variância (ANOVA). A ANOVA, proposta por Walpole et. al. (2009), utiliza a média dos resultados dos N experimentos para calcular a distância entre os grupos e verificar se essa diferença é muito próxima ou muito diferente.

Em cada um dos experimentos feitos na área da detecção de faces existem quatro variáveis resultantes que podem ser analisadas:

- Verdadeiro Positivo - Taxa de acerto de imagens de face que foram classificadas corretamente;
- Verdadeiro Negativo - Taxa de erro de imagens de face que foram classificadas incorretamente;
- Falso Positivo - Taxa de acerto de imagens de não face que foram classificadas corretamente;

- Falso Negativo - Taxa de acerto de imagens de não face que foram classificadas incorretamente.

Fez-se necessário o uso da média-f (*F-score*)⁹ representada na Equação 1, que convertia os valores de cada variável descrita acima para uma média única que foi utilizada no preenchimento da tabela 7 (Seção 5.7), que foi usada para os cálculos da ANOVA.

Equação 1

$$F = \frac{(1 + \beta) * \text{Verdadeiro}_{\text{Positivo}}}{((1 + \beta) * \text{Verdadeiro}_{\text{Positivo}} + \beta^2 * \text{Falso}_{\text{Negativo}} + \text{Falso}_{\text{Positivo}})}$$

Na Equação 1 o símbolo β representa o grau de importância dos valores. Como padrão foi adotado $\beta=1$.

Após a conversão de todos os resultados de todos os experimentos, finalmente o cálculo da ANOVA foi aplicado na tabela geral utilizando uma planilha eletrônica que gerou os resultados automaticamente. Com a análise dos resultados da ANOVA foi visto que os experimentos possuem grande diferença entre eles favorecendo a possibilidade de decidir qual o melhor método. A próxima seção irá detalhar todos os resultados obtidos.

⁹ <http://en.wikipedia.org/wiki/F-score>

5 - RESULTADOS

Dos vários resultados obtidos nesta pesquisa, foi decidido pelo detalhamento e análise apenas dos melhores resultados de cada experimento. Todos os valores dos resultados foram usados, porém seria inviável detalhamento de todas as 140 tabelas neste relatório. É possível verificar na Tabela 7 todas as médias *F-score* de todas as 140 tabelas. Para melhor entendimento, os resultados das análises dos experimentos globais e locais serão apresentados separadamente. Uma particularidade relevante que vale ser frisada diz respeito à sensibilidade das RNAs, para bons resultados serem obtidos utilizando RNA, fez-se necessário em cada experimento de cada conjunto de imagens a repetição do mesmo treinamento pelo menos dez vezes até encontrar o melhor treinamento comparado com o mesmo conjunto de teste. Ao obter a melhor classificação entre as repetições, o treinamento correspondente foi armazenado e seu valor assumido como principal para aquele experimento, descartando os demais treinamentos. Esse procedimento não foi necessário para o uso do classificador SVM.

Na Seção 5.1 serão apresentados os resultados dos métodos extraídos localmente. A Seção 5.2 apresentará de forma geral, os resultados dos métodos extraídos localmente. Em seguida, na Seção 5.3 detalhará os resultados utilizando como método de extração de característica o histograma de cinza 7x7 pixels e 3x3 pixels contidos na Seção 5.2. O mesmo ocorre na Seção 5.4, porém, detalhando os resultados do LBP 7x7 pixels e 3x3 pixels contidos na Seção 5.2. A Seção 5.5 compara os resultados obtidos utilizando como métodos de extração de característica o histograma de cinza 7x7 pixels e LBP 7x7 pixels contidos na Seção 5.2. Na Seção 5.6 será detalhado os resultados obtidos utilizando como método de extração de característica o histograma 3x3 pixels e LBP 3x3 pixels contidos na Seção 5.2.

5.1 - CARACTERÍSTICAS EXTRAÍDAS GLOBALMENTE

Inicialmente serão mostrados os experimentos feitos utilizando o procedimento de extração de característica global nos seguintes métodos: valores dos Pixels e PCA como é possível verificar na Tabela 1.

Tabela 1 - Resultados dos experimentos globais.

Taxa de Acerto em %	SVM		RNA	
	Face	Não Face	Face	Não Face
Valores Dos Pixels	71%	99%	87%	98%
PCA	92%	98%	89%	97%

Para melhor entendimento da Tabela 1 pode-se ressaltar que cada linha refere-se à taxa de acerto por parte do classificador correspondente à coluna. Como é possível verificar na Tabela 1, a taxa de acerto utilizando os valores dos pixels foi melhor classificada utilizando RNA, uma diferença significativa de **16%**. Por outro lado o PCA obteve melhor resultado geral utilizando SVM como classificador, pode-se concluir que o método de extração global utilizando PCA é mais eficiente em termos de processamento por utilizar 5 vezes menos quantidade de características do que as utilizadas nos experimentos feitos com os valores dos pixels. O PCA também obteve melhor classificação utilizando uma RNA do que os valores dos pixels, o que reforça seu poder de extração dos principais conteúdos das imagens. Mesmo com valores inferiores utilizando RNA em comparação com a combinação feita utilizando SVM como classificador, é notória a estabilidade da RNA em comparação com a SVM. Enquanto a RNA obteve uma diferença de **2%** nas taxas de acerto para face e **1%** nas taxas de acerto de não face entre os métodos utilizados, com SVM a diferença foi bastante significativa nas taxas de acerto de faces, ficando a diferença de **21%** o que é totalmente inaceitável.

5.2 - CARACTERÍSTICAS EXTRAÍDAS LOCALMENTE

Aqui serão apresentados os resultados dos métodos locais de extração de características, o modelo de tabela usado a seguir será idêntico ao modelo utilizado anteriormente.

Vale lembrar que alguns pontos neste método de extração devem ser levados em consideração. Como foi dito anteriormente, alguns métodos de extração de características possuem particularidades que seu uso aplicado na forma de extração global é totalmente inútil, tendo em vista a forma de atuação de seus algoritmos. No caso do LBP que utiliza no fim do seu processo a quantização de histogramas aplicada a nova matriz de valores obtida, com isso se faz necessária a diminuição em regiões da atuação do algoritmo, o que aperfeiçoa

a quantidade de *bins* do LBP melhorando sua precisão do padrão obtido, assim como na utilização do Histograma de Cinza. De todos os testes que necessitaram da utilização de *bins* para seu funcionamento, os experimentos foram repetidos exaustivamente em busca da melhor quantidade de *bins* para o extrator de característica e classificador correspondente.

Da mesma forma, o método de extração de característica utilizando os valores dos pixels, por utilizar todos os pixels da imagem, a divisão dela seria desnecessária. A única alternativa seria, ao invés de usar todos os pixels, utilizar pixels alternados da imagem, mas esse procedimento foi descartado e não teria como característica o método local, já que os valores seriam obtidos em sequências contínuas mesmo alternando entre si.

Outra particularidade ficou por conta da utilização do método PCA aplicado localmente, pois em todos os outros processos as imagens estavam dimensionadas em 21x21 pixels. Como o PCA possui a característica de generalização dos dados presentes nas imagens, seria desnecessária a divisão da imagem em muitas regiões, logo, foi decidido que a imagem deveria ser dividida em 4 regiões iguais. Para isso ser possível, deveria existir um redimensionamento da imagem para 20x20 pixels ou 22x22 pixels, números vizinhos do tamanho usado (21x21 pixels). Neste caso, foi decidido que a imagem deveria ser redimensionada para 20x20 pixels exclusivamente para o uso do PCA localmente em 4 regiões de 10x10 pixels cada.

Tabela 2 - Resultados dos experimentos locais.

Taxa de Acerto em %	SVM		RNA	
	Face	Não Face	Face	Não Face
Histograma de Cinza 7x7	62%	98%	74%	96%
Histograma de Cinza 3x3	87%	98%	98%	94%
PCA 10x10	70%	99%	81%	94%
LBP 7x7	66%	84%	99%	86%
LBP 3x3	96%	89%	99%	93%

Na Tabela 2, na coluna da esquerda onde há o nome do método utilizado na extração da característica da imagem, existem como rótulos 7x7, 10x10 ou 3x3, que indicam a quantidade de pixels de cada região que o método foi aplicado na imagem.

É possível verificar que a aplicação do PCA localmente teve uma queda bastante significativa nas taxas de acerto em comparação ao método de extração global, como visto na Tabela 1. Essa queda foi de **22%** para o acerto de faces no uso do classificador SVM e **8%** entre as taxas de acerto de faces usando RNA. Como dito anteriormente, a RNA tem uma característica mais regular em seus resultados, embora seja sensível à variação dos dados. As diferenças nas taxas de acerto para não face se manteve regular com pouca variação nos dois classificadores utilizando como método o PCA.

Uma característica importante de destacar é a variação significativa entre os mesmos métodos ao dividir a imagem em regiões com 7x7 ou 3x3 pixels. De todos os métodos utilizados localmente, é possível verificar que todas as taxas de acerto na classificação de faces usando RNA foram superiores às taxas usando como classificador a SVM. As taxas de acerto para não face entre os classificadores foram variadas, com no máximo **4%** de diferença entre o método utilizado combinado com os classificadores. Essa diferença utilizando métodos globais foi de **1%**.

5.3 - ANÁLISE DO RESULTADO UTILIZANDO HISTOGRAMA DE CINZA

Analisando a variação dos métodos isoladamente alterando o número de regiões, é possível confirmar que quanto mais regiões a imagem possuir melhor a taxa de acerto do classificador, como mostra a Tabela 3. Utilizando histogramas de cinza 7x7 pixels (9 regiões), a diferença das taxas de acerto entre os classificadores são de **12%** para face, com a maior taxa de acerto usando RNA e **2%** a diferença de classificação para não face, sendo **98%** a melhor taxa usando SVM. Em histograma de cinza (7x7 pixels), as melhores quantidades de *bins* para os experimentos realizados foram de 9 *bins* para SVM como classificador e 15 *bins* para o uso da RNA.

Tabela 3 - Análise isolada dos resultados usando histograma de cinza.

Taxa de Acerto em %	SVM		RNA	
	Face	Não Face	Face	Não Face
Histograma de Cinza 7x7	62%	98%	74%	96%
Histograma de Cinza 3x3	87%	98%	98%	94%

Ao dividir a imagem em mais regiões utilizando histogramas de cinza, é visível o aumento nas taxas de acerto dos classificadores com relação às imagens de faces usando

como método de extração de característica o histograma de cinza 3x3 pixels (49 regiões) comparadas com as taxas de acerto utilizando histograma de cinza 7x7 pixels (9 regiões), como mostra a Tabela 3. Enquanto a diferença nas taxas de acerto utilizando SVM como classificador foi de **25%** para imagens de face e para as imagens de não face ficando inalterada, a diferença entre as taxas de acerto do classificador RNA foi de **24%** a mais do que histogramas de cinza 7x7 pixels (9 regiões). A taxa de acerto para não face usando RNA caiu **2%**. A quantidade de *bins* utilizados nesses procedimentos foram de 8 *bins* para a combinação com o classificador SVM e 9 *bins* para o uso do RNA como classificador.

5.4 - ANÁLISE DOS RESULTADOS UTILIZANDO LBP

Na Tabela 4 é possível verificar a diferença nas taxas de acerto por parte dos classificadores SVM e RNA utilizando como métodos de extração de características o LBP 7x7 pixels (9 regiões) e LBP 3x3 pixels (49 regiões). A diferença entre as taxas de acerto das imagens com face chega a ser de **33%** entre os classificadores, sendo a maior das taxa do classificador RNA. Essa diferença é minimizada ao utilizar o LBP 3x3 pixels (49 regiões), em que ao ser combinado com o classificador SVM a taxa de acerto subiu **33%** em comparação com o mesmo classificador utilizando LBP 7x7 pixels (9 regiões). A taxa de acerto utilizando RNA se manteve constante para imagens de face, permanecendo **3%** melhor que a taxa de acerto utilizando SVM e melhora de **4%** na taxa de acerto de imagens de não face.

Tabela 4 - Análise isolada dos resultados usando LBP.

Taxa de Acerto em %	SVM		RNA	
	Face	Não Face	Face	Não Face
LBP 7x7	66%	84%	99%	86%
LBP 3x3	96%	89%	99%	93%

As quantidades de *bins* utilizados para o método LBP 7x7 pixels utilizando SVM foram de 7 *bins* e 8 *bins* para a combinação com o classificador RNA. Essas mesmas quantidades de *bins* foram utilizadas para o método LBP 3x3 pixels.

5.5 - ANÁLISE DOS RESULTADOS EM REGIÕES COM 7X7 PIXELS

Nesta seção serão analisados isoladamente os piores resultados dos dois métodos de extração de características. Utilizando regiões de 7x7 pixels são verificáveis as seguintes

distinções: ambos os classificadores tiveram melhor taxa de acerto para imagens de não face utilizando histogramas de cinza, ao contrário das imagens de face, onde o LBP obteve melhor taxa de acerto; o histograma de cinza teve taxa de acerto para imagem de não face de **98%** utilizando o SVM como classificador, porém a taxa de acerto para face foi de **62%**, o que é significativamente baixa; ao utilizar o LBP combinado com RNA a taxa de acerto foi de **99%**, mas a taxa de acerto para não face foi de **86%**, como mostra a Tabela 5. Com isso, surge uma imparcialidade na escolha de qual o melhor método. Essa afirmação só é possível com análise estatística de todos os resultados, daí a importância dos dez grupos de imagem para a realização de vários experimentos. Esta análise será apresentada na seção 5.7.

Tabela 5 - Análise isolada dos resultados obtidos em regiões com 7x7 pixels.

Taxa de Acerto em %	SVM		RNA	
	Face	Não Face	Face	Não Face
Histograma de Cinza 7x7	62%	98%	74%	96%
LBP 7x7	66%	84%	99%	86%

5.6- ANÁLISE DOS RESULTADOS EM REGIÕES COM 3X3 PIXELS

Serão analisados isoladamente nesta seção os resultados utilizando a aplicação dos métodos histogramas de cinza e LBP em regiões com 3x3 pixels. Esses resultados são notavelmente vistos como os melhores de ambos os métodos de extração de características entre os demais, como é possível verificar na Tabela 6.

Tabela 6 - Análise isolada dos resultados obtidos em regiões com 3x3 pixels.

Taxa de Acerto em %	SVM		RNA	
	Face	Não Face	Face	Não Face
Histograma de Cinza 3x3	87%	98%	98%	94%
LBP 3x3	96%	89%	99%	93%

Na Tabela 6, é possível verificar com maior clareza que a imagem ao ser dividida em mais regiões as taxas de acerto por parte dos classificadores têm aumento bastante significativos. Ao comparar as Tabelas 6 e 5, é visto que o método de histograma de cinza 3x3 tem um aumento na taxa de acerto para imagens com face de aproximadamente **25%** em ambos os classificadores. No uso do método de LBP 3x3 a melhora na taxa de acerto de face

se deu apenas no classificador SVM. O classificador RNA se manteve estável com melhoras de **7%** na taxa de acerto de imagens de não face.

Todas as tabelas apresentadas nesta seção contêm os melhores resultados de cada método de todos os dez conjuntos de experimentos. Inicialmente, ao analisar os resultados expostos até o momento, poderia ser atribuído que o método de extração de característica de histograma de cinza 3x3 pixels combinado com o classificador RNA possuiria o melhor resultado nas taxas de acerto comparado com os demais. Com isso, vale lembrar que a análise foi feita de modo geral, utilizando todos os dez experimentos de todos os métodos. Os resultados apresentados até o momento são os melhores resultados de cada método conseguidos em um dos dez experimentos realizados.

A seção a seguir tratará estatisticamente com todos os resultados de todos os dez experimentos realizados utilizando todos os métodos de extração de características, permitindo assim classificar os melhores ou piores métodos.

5.7 - AVALIAÇÃO DE TODOS OS RESULTADOS USANDO ANÁLISE DE VARIÂNCIA (ANOVA)

Sabendo-se todos os resultados de todos os experimentos feitos, estes foram convertidos utilizando a Equação 1 apresentada na Seção 4.3. Os valores de cada resultado foram usados para o preenchimento da Tabela 7, que compõe a tabela principal onde os dados serão analisados aplicando ANOVA. Na Tabela 7, a primeira coluna da esquerda representa os nomes de cada método de extração de característica referente às linhas. A coluna logo a seguir dos nomes dos métodos fazem referência aos tipos de classificadores utilizados. A primeira linha da tabela corresponde a quantidade de experimentos, como dito anteriormente. A última coluna corresponde a média F-score do melhor resultado do método correspondente. Ao analisar a Tabela 7 fica clara que a média F-score oferece um grau de importância nas taxas de acerto das imagens de face e não face por igual. Desta forma, é possível determinar e classificar quais dos métodos tiveram melhores resultados. Aqui, os valores foram distribuídos com precisão de 4 dígitos, mas nos cálculos foram usados precisão de 9 dígitos.

Tabela 7 - Distribuição dos valores convertidos usando *F-score* para análise de variância.

	Experimento	1	2	3	4	5	6	7	8	9	10	Maior Valor
Valores dos Pixels	SVM	0,8291	0,8127	0,8285	0,8256	0,8201	0,8214	0,8176	0,8126	0,8198	0,8201	0,8291
	RNA	0,9100	0,8958	0,9024	0,9055	0,9092	0,9101	0,8991	0,8962	0,8995	0,9091	0,9101
Histograma de Cinza 7x7	SVM	0,7505	0,7532	0,7465	0,7577	0,7502	0,7562	0,7457	0,7467	0,7500	0,7458	0,7577
	RNA	0,8200	0,8315	0,8222	0,8311	0,8216	0,8216	0,8301	0,8198	0,8196	0,8256	0,8315
PCA Global	SVM	0,9385	0,9386	0,9399	0,9430	0,9399	0,9401	0,9412	0,9406	0,9424	0,9402	0,9430
	RNA	0,9101	0,9124	0,9112	0,9167	0,9151	0,9145	0,9111	0,9085	0,9091	0,9089	0,9167
LBP 7x7	SVM	0,7212	0,7167	0,7212	0,7219	0,7165	0,7157	0,7189	0,7211	0,7195	0,7200	0,7219
	RNA	0,9315	0,9341	0,9301	0,9299	0,9331	0,9315	0,9322	0,9299	0,9301	0,9299	0,9341
Histograma de Cinza 3x3	SVM	0,9112	0,9096	0,9102	0,9096	0,9151	0,9175	0,9197	0,9176	0,9161	0,9131	0,9197
	RNA	0,9401	0,9401	0,9414	0,9445	0,9451	0,9453	0,9451	0,9441	0,9450	0,9450	0,9453
PCA 10x10	SVM	0,8177	0,8162	0,8175	0,8178	0,8088	0,8101	0,8164	0,8170	0,8085	0,8120	0,8178
	RNA	0,8512	0,8601	0,8595	0,8518	0,8581	0,8655	0,8610	0,8663	0,8602	0,8612	0,8663
LBP 3x3	SVM	0,9221	0,9212	0,9240	0,9240	0,9231	0,9231	0,9203	0,9222	0,9200	0,9200	0,9240
	RNA	0,9531	0,9514	0,9561	0,9554	0,9553	0,9524	0,9513	0,9541	0,9560	0,9561	0,9561

A Tabela 8 apresenta o resumo dos cálculos feitos na Tabela 7 utilizando a precisão de 9 casas decimais. A coluna da esquerda apresenta os nomes dos métodos, a coluna seguinte contém o número de repetições dos valores, nesse caso o valor 10 representa que os métodos foram aplicados 10 vezes diferentes. As demais colunas apresentam a soma geral de todos os 10 experimentos do método correspondente, a média das somas e a variância dos valores dos métodos correspondentes.

A variância apresentada na Tabela 8 representa o quanto os valores contidos na Tabela 7 entre cada método de extração de característica variaram entre eles com o mesmo classificador. Ou seja, a distância dos valores contidos em cada linha da Tabela 7, entre o pior resultado e o melhor resultado.

Quanto maior for essa variância, maior é a distancia do pior resultado para o melhor resultado no método correspondente, sendo assim podemos considerar que o classificador se manteve instável dentro daquele método de extração de característica. Quanto menor for essa variância, mais estável foi o comportamento do classificador utilizando o método de extração correspondente. Sendo assim, pode-se considerar que o método utilizando os valores dos pixels combinado com o classificador RNA obteve maior variância, enquanto o método PCA combinado com o classificador SVM a menor variância.

Tabela 8 - Resumo da análise geral da variância.

Métodos	Contagem	Soma	Média	Variância
Valores dos pixels - SVM	10	8,207510256	0,820751026	3,33167E-05
Valores dos pixels - RNA	10	9,036980446	0,903698045	3,34562E-05
Histograma de cinza 7x7 - SVM	10	7,50253905	0,750253905	1,86171E-05
Histograma de cinza 7x7 - RNA	10	8,242947259	0,824294726	2,37701E-05
PCA - SVM	10	9,404328191	0,940432819	2,11492E-06
PCA - RNA	10	9,117707016	0,911770702	8,02853E-06
LBP 7x7 - SVM	10	7,192641628	0,719264163	5,04936E-06
LBP 7x7 - RNA	10	9,312368537	0,931236854	2,28452E-06
Histograma de cinza 3x3 - SVM	10	9,139806102	0,91398061	1,38269E-05
Histograma de cinza 3x3 - RNA	10	9,435803645	0,943580365	4,63409E-06
PCA 10x10 - SVM	10	8,142002671	0,814200267	1,50348E-05
PCA 10x10 - RNA	10	8,595087788	0,859508779	2,4089E-05
LBP 3x3 - SVM	10	9,220221215	0,922022122	2,4374E-06
LBP 3x3 - RNA	10	9,541056036	0,954105604	3,79131E-06

Tabela 9 - Resultado da análise de variância ANOVA.

Fonte da variação	SQ	GL	MQ	F	valor-P	F-crítico
Entre grupos	0,734575448	13	0,056505804	4153,726011	4,6851E-159	1,798584005
Dentro dos grupos	0,001714059	126	1,36036E-05			
Total	0,736289507	139				

Na Tabela 9, são apresentados os valores significativos dos cálculos da análise de variância. Na coluna da esquerda estão os rótulos das linhas. Na coluna com rotulo **SQ** estão os valores das somas dos quadrados. O **GL** significa o grau de liberdade dos grupos. **MQ** é a média dos quadrados, é obtido dividindo o **SQ** pelo **GL**. O valor de **F** é encontrado na divisão do **MQ** Entre grupos pelo **MQ** Dentro dos grupos. O **valor-p** é a probabilidade referente à hipótese nula ser validada, se o **valor-p** for menor que o limiar usado, neste caso 0.05, então a hipótese nula é descartada. O valor contido na coluna **valor-p**, representa o número que falta para a taxa de 100% da validação dos dados. **F-crítico** é um valor tabelado baseado nos valores do **GL**.

Uma das aplicações que a ANOVA pode ser adotada é a de testar hipóteses. É atribuída hipótese nula (H_0) que assume o pressuposto de que todas as variáveis não são significativamente diferentes e a hipótese H_1 , que assume o pressuposto de que todas as variáveis são significativamente diferentes.

No caso do cálculo da ANOVA aplicado nas médias dos dados *F-score*, é possível verificar na Tabela 9 que o valor de F foi muito superior ao F-Crítico, então se pode adotar a hipótese (H_1) que infere dizer que as médias do *F-score* de todos os métodos utilizando diferentes classificadores são diferentes. Neste caso, pode-se afirmar que, de acordo com a análise ANOVA, a validade da hipótese H_1 se deu pela diferença das médias entre os grupos (0,734575448) e não pelas médias dentro do grupo (0,001714059), que foi muito pequena.

De acordo com a ANOVA, esta pesquisa validou a hipótese H_1 , pois a análise dos dados comprovaram as diferenças entre as classes, ou seja, os resultados obtidos utilizando vários métodos de extração de características combinados com dois tipos de classificadores são significativamente diferentes.

Logo, é possível qualificar, de forma concreta, quais os melhores e piores métodos de extração de características com os respectivos classificadores utilizando os valores obtidos em seus experimentos, de acordo com as médias *F-score* apresentadas pela Tabela 8.

A Tabela 10 apresenta um *ranking* em ordem decrescente dos métodos de extração de características combinados com os classificadores SVM e RNA, utilizando as médias *F-score* dos experimentos realizados.

Tabela 10 - *Ranking* dos métodos e classificadores pelas médias *F-score*.

Posição	Método	Regiões	Classificador	Média
14	LBP	7x7	SVM	0,719264163
13	Histograma de cinza	7x7	SVM	0,750253905
12	PCA	10x10	SVM	0,814200267
11	Valores dos pixels	Global	SVM	0,820751026
10	Histograma de cinza	7x7	RNA	0,824294726
9	PCA	10x10	RNA	0,859508779
8	Valores dos pixels	Global	RNA	0,903698045
7	PCA	Global	RNA	0,911770702
6	Histograma de cinza	3x3	SVM	0,913980610
5	LBP	3x3	SVM	0,922022122
4	LBP	7x7	RNA	0,931236854
3	PCA	Global	SVM	0,940432819
2	Histograma de cinza	3x3	RNA	0,943580365
1	LBP	3x3	RNA	0,954105604

6 - CONCLUSÃO

Este trabalho analisou influência das imagens com vários níveis de iluminação na detecção de faces. Vários métodos de extração de características foram aplicados em dois tipos de classificadores em busca de identificar qual a melhor combinação para entre eles que pudesse oferecer uma solução para o problema abordado.

Foi verificado que, nem todos os métodos de extração de características podem ser aplicados da mesma forma, a exemplo do PCA, que os valores dos seus experimentos realizados globalmente foram bastante diferentes quando realizados de modo local. Foi possível analisar, na mesma proporção, que os métodos de extração que dividiram as imagens em várias regiões obtiveram resultados melhores que os métodos que as dividiram em menos regiões, ao contrário do que aconteceu com o PCA.

Observando essas distinções, pode-se concluir que cada método possui uma característica importante que deve ser usada especificamente para contornar um determinado problema.

Outro fator importante se deu pelas taxas de acerto usando como método de extração de característica os valores dos pixels, que foram muito baixas utilizando toda a informação da imagem se comparada com o PCA, que utilizou a proporção de 20% das informações da imagem, o que leva a conclusão que para imagens com variância de iluminação e conseqüentemente, sobreamento, utilizar toda a informação da imagem não é adequada, pois seu conteúdo redundante afeta a classificação.

A aplicação dos resultados na Equação 1 (F-score) foi importante em todo o procedimento de análise dos resultados, pois sua fórmula dava devida importância para a taxa de erro por parte do classificador, diferenciando os grupos e as classes entre cada método. Além disso, o uso de F-score permite obter um número único para avaliar um classificador.

A aplicação da ANOVA foi importante no processo, tendo em vista que, houve a análise que todos os métodos de extração de característica eram diferentes entre si. A ANOVA confirmou e validou estatisticamente a diferença entre as classes internas (os dez experimentos de um mesmo método) e a diferença entre todas as classes de grupos (todos os

métodos de extração possuem resultados diferentes entre si). Essa afirmação é possível, pois de acordo com a Tabela 9, o valor de F é muito superior ao valor do F-crítico.

Partindo da informação de que o experimento é estatisticamente válido, é possível classificar os métodos de extração de características de acordo com seus resultados nas médias F-score e quais as combinações que obtiveram os melhores valores nas taxas de acerto do classificador correspondente. A Tabela 10 é resultado da coluna das médias contida na Tabela 8. Por esses valores foi possível criar um *ranking* dos melhores resultados obtidos.

Pelos resultados obtidos nos experimentos, é possível concluir que o método de extração de característica LBP dividindo a imagem em regiões com dimensões de 3x3 pixels combinado com uma RNA como classificador dos padrões adquiridos, obteve o melhor resultado nas taxas de acerto em imagens com condições de iluminação variadas.

Com a segunda melhor taxa de acerto, com valor próximo do primeiro, ficou o método de extração que utilizou histograma de cinza, também utilizando regiões com dimensões de 3x3 pixels e utilizando como classificador dos padrões adquiridos uma RNA.

Sendo assim, é possível concluir que, para essas condições de imagens, quanto maior for a quantidade de regiões, melhor será a taxa de acerto. Podemos inferir que o melhor classificador foi a RNA pelo fato do seu poder de adequação a qualquer tipo de dado inserido e sua automação da organização dentro dela.

Dos resultados utilizando SVM como classificador, a melhor taxa de acerto obtida foi combinada com PCA aplicado globalmente como método de extração de característica que ficou na terceira colocação no *ranking*.

Até o momento o que foi visto nas referências bibliográficas foi que muitos pesquisadores, na área da detecção de faces, passam por cima do problema de iluminação e sombreamento da face sem nenhum tratamento adequado ou o tratam com algum método de extração de característica combinado com outro método, que funciona para alguns casos. Com essas observações é possível concluir que este trabalho tem uma grande contribuição na área da detecção de faces, pois sua abordagem foi minuciosa, propondo a melhor combinação entre os métodos de extração de característica e classificadores para contornar ou minimizar o problema de iluminação e sombreamento em faces.

REFERÊNCIAS BIBLIOGRÁFICAS

BAIOCO, R. L.; SALOMÃO, J. M. Detecção De Faces Em Locais Públicos E Com Aglomeração De Pessoas. Instituto Federal de Educação Ciência e Tecnologia do Espírito Santo, 2010.

CARVALHO, F. J. S. ; TAVARES, J. M. R. S. Detecção de Faces em Imagens baseada na Identificação da Pele e dos Olhos, Revista de Estudos Politécnicos, Tékhné, Vol VI, nº9, 2008.

DIAS, C. G. ; RADONSKY, V. Desenvolvimento de um sistema de auxílio ao diagnóstico em pediatria com o uso de redes neurais artificiais, Exacta, abril, volume 1, Centro Universitário Nove de Julho; São Paulo, Brasil; p. 89-95; 2003.

FREUND, Y. ; SCHAPIRE R. E. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55, 1997.

GEORGHIADES, A. S. ; BELHUMEUR, P.N. ; KRIEGMAN, D. J. From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. IEEE Trans. Pattern Anal. Mach. Intelligence, 2001. Vol. 3, n.6, p. 643-660.

GORBENKO, A. ; POPOV, V. On Face Detection from Compressed Video Streams, Applied Mathematical Sciences, Vol. 6, no. 96, 4763 - 4766, may, 2012.

HARALICK, R. M. ; SHANMUGAM, K. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics. v. 3, SMC-3, n. 6, p. 610-621, nov. 1973.

HAYKIN, S. Neural Networks: A Comprehensive Foundation, 2nd Edition. Prentice Hall, 1998. BAL h 01:1 1.Ex.

JAIN, R. ; KASTURI, R. ; SCHUNK B. G. Published by McGraw-Hill, Inc. 1995.

LORENA, A. C. ; CARVALHO, A. C. P. L. Introdução as Máquinas de Vetores Suporte. São Carlos - SP, Abril 2003.

MARCEL, S. ; RODRIGUEZ, Y. ; HEUSCH, G. On the recent use of Local Binary Patterns for Face Authentication, International Journal of Image and Video Processing, Special Issue on Facial Image Processing, 1-9, 2007.

MONARD, M. C. ; BARANAUSKAS, J. A. Sistemas Inteligentes: Fundamentos e Aplicações. Capítulo 4 - Conceito sobre aprendizagem de máquina, p. 89–114. Editora Manol, 2003.

NIXON, M. ; AQUADO, A. S. Feature Extraction & Image Processing for Computer Vision, Third Edition, Elsevier, 2012.

OJALA, T. ; PIETIKÄINEN, M. ; HARDWOOD, D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition, 29(1):51-59, 1996.

OJALA, T. ; PIETIKAINEN, M. ; MAENPAA, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 971–987, 2002.

PAPAGEORGIOU, O. M. ; POGGIO, T. A general framework for object detection, Sixth International Conference on Computer Vision, p. 555 - 562, 1998.

PEARSON, K. On Lines and Planes of Closest Fit to Systems of Points in Space. Philosophical Magazine 2 (11): p. 559–572. 1901.

PEREIRA, E. T. Detecção Automática de Faces: Ferramenta para Combate ao Crime. Revista Farol Digital, Paraiba, p. 40 - 41, 01 jun. 2008.

PEREIRA, E. T. ; GOMES, H. M. ; CARVALHO, J. M. ; Integral Local Binary Patterns: a Novel Approach Suitable for Texture-Based Object Detection Tasks. In: 23rd SIBGRAPI Conference on Graphics, Patterns and Images, 2010, Gramado - RS. The SIBGRAPI Proceedings. Los Alamitos, CA: IEEE Computer Society Press, 2010.

PEREIRA, E. T. ; GOMES, H. M. ; MOURA, E. S. ; CARVALHO, J. ; ZHANG, T. Investigation of Local and Global Features for Face Detection. IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing. CIMSIVP, Paris, 2011.

PHAM, D. T. ; ALCOCK, R. J. Automated grading and defect detection: a review. Forest Products Journal, v. 48, n. 4, p. 34-42, 1998.

PORTUGAL, MARCELO S. ; FERNANDES, LUIZ G. L. Redes Neurais Artificiais e Previsão de Séries Econômicas: uma introdução, Nova Econômica, Vol. 6, n.1, 1996.

PRADO, J. L. ; GONZAGA, A. Detecção de Faces Humanas através da Proporção Áurea, VIII Workshop de Visão Computacional 2012.

QUINTANS, M. ; CARNEVALI, F. ; ESCOLA, J. Implementação de um software para reconhecimento facial utilizando HaarCascade através da OpenCV. I ERI-RJ – I Escola Regional de Informática do Rio de Janeiro Instituto de Matemática, p.72-75. 2010.

RIBEIRO, F. M. L. ; ARAUJO, G. M. ; SILVA, E. A. B. ; OLIVEIRA, J. F. L. ; GOLDENSTEIN, S. K. Detecção de Pontos Fiduciais sobre a Face em Tempo Real, xxx simpósio brasileiro de telecomunicações, Brasília, 2012.

RODRIGUEZ, Y. Face Detection and Verification using Local Binary Patterns. Phd thesis, Ecole Polytechnique Federale de Lausanne, 2006.

SILVA, A. C. S. ; ALMEIDA, T. S. ; BATISTA, E. DA C. ; REIS, E. DE J. C. ; REIS, ARTUR B. S. Algoritmo para detecção de boca em faces humanas usando matriz de coocorrência e svm, Pró-Reitoria de Pesquisa e Pós-Graduação (PPPG) da Universidade Federal do Maranhão (UFMA), Vol. 19: n. especial, 2012.

TAHIM, ANDRÉ P. N. Máquinas de vetores de suporte (svm), Março, 2010. Disponível em: www.andretahim.com/publicacoes. Acessado em: outubro 2012.

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

VASCONCELOS, S. Análise de Componentes Principais. Disponível em: www.ic.uff.br/~aconci/PCA-ACP.pdf. Acessado em: outubro 2012.

VELOSO, L. R. ; CARVALHO, J. M. ; CAVALVANTI, C. S. V. C. ; MOURA, E. S. ; COUTINHO, F. L. ; GOMES, H. M. Neural Network Classification of Photogenic Facial Expressions Based on Fiducial Points and Gabor Features. PSIVT 2007, p. 166-179.

VIOLA, P. ; JONES, M. Rapid object detection using a boosted cascade of simple features, Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, I-511-I-518, 2001a.

VIOLA, P. ; JONES, M. Robust real-time object detection. In Second International Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing, and Sampling, 2001b.

WALPOLE, R. E. ; MYERS, R. H. ; MYERS, S. L. ; YE, K. Probabilidade e estatística para engenharia e ciências, 8ª edição, Pearson/Prentice Hall, 2009.