



Universidade Federal da Bahia
Instituto de Matemática / Escola Politécnica

Programa de Pós-Graduação em Mecatrônica

**COHAWES: UM NOVO DESCRITOR
APLICADO EM RECONHECIMENTO FACIAL**

Leandro Costa Souza

DISSERTAÇÃO DE MESTRADO

Salvador
18 de dezembro de 2017

LEANDRO COSTA SOUZA

**COHAWES: UM NOVO DESCRITOR APLICADO EM
RECONHECIMENTO FACIAL**

Esta Dissertação de Mestrado foi apresentada ao Programa de Pós-Graduação em Mecatrônica da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Mestre em Mecatrônica.

Orientador: Prof. Dr. Luciano Rebouças de Oliveira

Salvador
18 de dezembro de 2017

Sistema de Bibliotecas - UFBA

Souza, Leandro Costa.

COHAWES: Um novo descritor aplicado em reconhecimento facial /
Leandro Costa Souza – Salvador, 2017.

68p.: il.

Orientador: Prof. Dr. Prof. Dr. Luciano Rebouças de Oliveira.

Dissertação (Mestrado) – Universidade Federal da Bahia, Instituto de
Matemática / Escola Politécnica, 2017.

1. Extração de características. 2. Haar wavelet. 3. Correlação. 4.
Reconhecimento facial. 5. Descritor. I. Oliveira, Luciano Rebouças de
. II. Universidade Federal da Bahia. Instituto de Matemática / Escola
Politécnica. III Título.

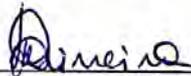
TERMO DE APROVAÇÃO

LEANDRO COSTA SOUZA

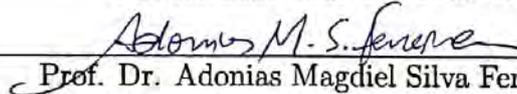
COHAWES: UM NOVO DESCRITOR APLICADO EM RECONHECIMENTO FACIAL

Esta Dissertação de Mestrado foi julgada adequada à obtenção do título de Mestre em Mecatrônica e aprovada em sua forma final pelo Programa de Pós-Graduação em Mecatrônica da Universidade Federal da Bahia.

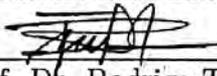
Salvador, 18 de Dezembro de 2017



Prof. Dr. Luciano Rebouças de Oliveira
Universidade Federal da Bahia



Prof. Dr. Adonias Magdiel Silva Ferreira
Universidade Federal da Bahia



Prof. Dr. Rodrigo Tripodi Calumby
Universidade Estadual de Feira de Santana

Dedico este trabalho os meus pais, irmãos, minha esposa Luane Andrade Barreto e a toda minha família que me apoiaram e torceram muito por esta vitória. Aos amigos e colegas, pelo incentivo. Dedico especialmente ao meu filho Gabriel Andrade Souza que antecipou nossos planos para se fazer presente, mesmo que no ventre da mãe, na defesa desta dissertação.

AGRADECIMENTOS

Agradeço imensamente ao meu professor Dr. Luciano Rebouças de Oliveira que direcionou, motivou e acreditou, mais até do que eu mesmo, na conclusão deste trabalho. Agradeço também a todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigado.

Tentar e falhar é, pelo menos, aprender. Não chegar a tentar é sofrer a inestimável perda do que poderia ter sido.

—GERALDO EUSTÁQUIO DE SOUZA

RESUMO

O reconhecimento automático de faces em imagens tem o objetivo de identificar indivíduos a partir de algumas imagens previamente cadastradas em uma base de dados. A estratégia de reconhecimento deve ser robusta o suficiente para contornar os problemas inerentes a tal tarefa, tais como: baixa qualidade na imagem, mudanças na iluminação ou mesmo alterações na expressão facial, por exemplo. Para atacar tais problemas, métodos para representação robusta de objetos em imagem são de fundamental importância. Com o propósito de contribuir com uma nova representação de imagem para auxiliar no processo de reconhecimento de faces, este trabalho introduz um novo descritor de imagem, denominado Correlation of Haar wavelet Sums (COHAWES). Este novo descritor calcula a correlação em regiões da imagem com base nos somatórios dos gradientes verticais e horizontais de filtros haar wavelets. O descritor proposto é avaliado com a ajuda dos classificadores *nearest neighbor* (NN), *locally-constrained linear coding* (LLC), *k-nearest neighbor* com *sparse representation classifier* (KNN-SRC), *linear regression classification* (LRC), minimização ortonormal ℓ_2 (MO-L2), *support-vector machines* (SVM), *linearly approximated sparse representation-based classification* (LASRC), *sparse representation classifier* com *gradient projection for sparse reconstruction* (SRC-GPSR), e *sparse representation classifier with homotopy* (SRC-Homotopy). O COHAWES foi comparado a outros descritores, tais como: dual-cross patterns (DCP), histogram of oriented gradients (HOG), gabor wavelets, local binary patterns (LBP), um vetor combinando de HOG, Gabor wavelets e LBP (HOG+GABOR+LBP) e com o COHAWES-COV que é uma modificação do COHAWES utilizando covariância no lugar da correlação, utilizando os classificadores previamente listados. Essas comparações foram realizadas utilizando os data sets ORL, Yale, YaleB, FERET (FB, DUP1 e DUP2), PubFig83+LFW, tendo o COHAWES apresentado resultados superiores na média da precisão geral e individualmente na maioria dos casos. Avaliações em outros domínios de imagem ainda necessitam ser realizadas, a fim de se analisar a aplicação do COHAWES em outras tarefas de representação.

Palavras-chave: extração de características, Haar wavelet, correlação, reconhecimento facial, descritor;

ABSTRACT

Automatic face recognition in images looking-for identification of some previously registered images in a database. The recognition strategy should be robust enough to overcome the problems inherent in this task, such as low image quality, changes in lighting or even change in facial expression, for example. To confront such problems, methods for robust representation of objects in image are of primary importance. In order to contribute with a new image representation to assist in the process of face identification, this work introduces a new image descriptor, called Correlation of Haar Wavelet Sums (COHAWES). This new descriptor calculates the correlation in regions of the image based on the sums of vertical and horizontal gradients of haar wavelets filters. This descriptor is applied in face recognition and evaluated with the help of the nearest neighbor (NN), locally-constrained linear coding (LLC), k-nearest neighbor with sparse representation classifier (KNN-SRC), linear regression classification (LRC), orthonormal minimization ℓ_2 (MO-L2), support-vector machines (SVM), linearly approximated sparse representation-based classification (LASRC), sparse representation classifier with gradient projection for sparse reconstruction (SRC-GPSR), and sparse representation classifier with homotopy (SRC-Homotopy). The performance in face recognition using COHAWES was compared to others, such as: dual-cross patterns (DCP), histogram of oriented gradients (HOG), Gabor wavelets, local binary patterns (LBP), a vector combining HOG, Gabor wavelets and LBP (HOG + GABOR + LBP) and COHAWES-COV, which is a modification of COHAWES using covariance instead of correlation, using the previously listed classifiers. These comparisons were performed using the ORL, Yale, YaleB, FERET, PubFig83 + LFW data sets, and COHAWES presented superior results on the mean of the overall accuracy and in most cases in the majority of cases. Evaluations in other image domains still need to be performed in order to analyze the application of COHAWES in other representation tasks.

Keywords: feature extraction, Haar wavelet, correlation, face recognition, descriptor;

SUMÁRIO

Capítulo 1—Introdução	1
1.1 Motivação	4
1.2 Objetivos	4
1.3 Contribuições	5
1.4 Mapa dos capítulos	5
Capítulo 2—Sistemas de reconhecimento facial	7
2.1 Definições gerais	8
2.2 Extratores de características	10
2.2.1 Gabor wavelets	11
2.2.2 <i>Local binary patterns</i>	14
2.2.3 <i>Histogram of oriented gradients</i>	15
2.2.4 <i>Dual-cross patterns</i>	16
2.2.5 SURF	17
2.2.6 Descritor de covariância	22
2.3 Classificadores	23
2.3.1 Classificação não-supervisionada	23
2.3.2 Classificação supervisionada	24
2.4 Data sets	26
2.4.1 Facial recognition technology	27
2.4.2 The IMM face database	28
2.4.3 The ORL database of faces	29
2.4.4 Yale data sets	29
2.4.5 PubFig83+LFW (PF83LFW)	30
2.5 Relação com o trabalho proposto	30
Capítulo 3—Correlation of Haar wavelet Sums	33
3.1 Visão geral do COHAWES	33
3.2 Descrevendo cada pixel da imagem	35
3.3 Divisão da imagem em regiões	38
3.4 Obtendo o vetor de características	38
3.4.1 Correlação dentro de cada região	39
3.4.2 Concatenação do vetor de características	40
3.5 Considerações finais	41

Capítulo 4—Avaliação experimental	43
4.1 Avaliação experimental do descritor inicial	44
4.1.1 Classificação por meio de métricas de distâncias	45
4.1.2 Classificação supervisionada	47
4.1.3 Seleção de subconjunto de características	48
4.1.4 Utilização de um descritor de textura como características correlacionadas	49
4.2 Avaliando melhor o COHAWES	50
4.2.1 Metodologia	51
4.2.2 Análise comparativa entre COHAWES e outros métodos	52
4.2.3 Análise de resultados	56
Capítulo 5—Discussão e conclusões	59
5.1 Data set de “universo aberto”	59
5.2 Tempo para computar do COHAWES	60
5.3 Significância do coeficiente de correlação	60
5.4 Dispersão de dados diferentes com mesma correlação	61
5.5 Benchmark para reconhecimento facial	61
Referências Bibliográficas	63

LISTA DE FIGURAS

2.1	Fluxo geral de um sistema de reconhecimento facial. (a) Imagem a ser identificada, (b) face detectada com pontos fiduciais, (c) imagem unicamente da face, (d) representação da face a partir de um vetor de características, (e) busca da face no banco de faces e (f) identificação da face propriamente dita.	8
2.2	(a) Estratégia de reconhecimento: o sistema busca a identidade de uma face-de-busca a partir de uma galeria com faces-alvo. (b) Estratégia de verificação: uma face-alvo reivindica uma identidade ao sistema, o qual responde que sim ou não, caso a face-alvo tenha acesso ou não ao referido sistema, respectivamente.	9
2.3	Espaço de representação de objetos não-face, face e dentro de face a separação entre faces de pessoas diferentes. Adaptado de (LI; JAIN, 2011).	10
2.4	Exemplos de abordagens para extração de características de uma imagem de face: (a) Apresenta a abordagem holística onde toda a face é considerada. (b) Apresenta a abordagem estrutural onde componentes da face ou regiões da imagem são consideradas	11
2.5	Representação dos grafos que descrevem duas imagens de face. Representação de uma face através de um grafo. Quando essas distorções estão dentro do valor de aceite as imagens são consideradas da mesma face. Figura extraída de (LADES et al., 1993).	12
2.6	Aplicação de núcleos de Gabor de forma holística na face. (a) 40 núcleos de Gabor obtidos por uma variação de 5 escalas diferentes e 8 orientações. (b) Convolução da imagem da face com cada um dos núcleos de Gabor definidos. (c) Junção da amostragem em baixa resolução das imagens geradas pela convolução da imagem da face com os núcleos de Gabor criando características para representação da face. Figura extraída de (LIU; WECHSLER, 2002)	13
2.7	Exemplo de deformação das malha utilizadas para calcular os pontos de convoluções com os filtros de Gabor. (a) Exibição da malha inicial (em amarelo) e forma média da face (em azul). (b) Forma padrão da face ajustada (verde) com a da imagem e malha sofrendo as mesmas deformações (amarelo) fazendo com que a posição dos pontos de convoluções com os filtros de Gabor tenham a mesma posição relativa às características de uma face em diferentes imagens. As figuras (c) e (d) mostram a posição dos pontos de convoluções com os filtros de Gabor em face frontal e rotacionada, respectivamente. Figura extraída de (CAMENT et al., 2015).	13

- 2.8 Cálculo do Local Binary Patterns. (a) Imagem Original. Ao lado da imagem original é mostrado um exemplo de uma vizinhança 3×3 . Todos os pixels que tiveram um valor maior, ou igual, ao pixel central receberá o valor 1 e os que tiverem valores menores receberá 0. A sequência de 0s e 1s é convertida para base 10 (b) Imagem após cada pixel ser substituído pelo LBP. 14
- 2.9 Extração do *Histogram of oriented gradients*. A imagem Original em escala de cinza é filtrada duas vezes, em x e em y gerando g_x e g_y ; com esses valores são calculados a magnitude g e a orientação θ ; a magnitude g é dividida em células de 8×8 e essas células são agrupadas em blocos sobrepostos de 2×2 (em vermelho e em verde); é feito o histograma da magnitude g agrupado por 9 *bins* da orientação θ ; a junção de todos os histogramas de todos os blocos compõem o HOG. 16
- 2.10 Representação do rosto usando DCP. A imagem de face normalizada é codificada pelos dois encoders cruzados, respectivamente. A concatenação dos histogramas regionais do código DCP forma a representação de face baseada em DCP. Figura extraída de (DING et al., 2016). 17
- 2.11 Da esquerda para a direita: Derivadas parciais gaussianas de segunda ordem (discretizadas e recortadas), com $\sigma = 1.2$, na direção y e direção xy , e aproximações usando filtros de caixa. As regiões cinza são iguais a zero. Figura extraída de (BAY; TUYTELAARS; GOOL, 2006). 18
- 2.12 Atribuição de orientação: uma janela de orientação deslizante de tamanho $\frac{\pi}{3}$ detecta a orientação dominante das Haar wavelets ponderadas pela Gaussiana em cada ponto de amostra dentro de uma vizinhança circular ao redor do ponto de interesse. Adaptado de (BAY et al., 2008). 19
- 2.13 Para construir o descritor é colocado uma grade quadrada orientada dividida em sub-regiões 4×4 sobre o ponto de interesse (centro). Para cada sub-região, as respostas das Haar wavelets são calculadas a partir de amostras 5×5 . Para cada região, são coletadas as somas d_x , $|d_x|$, d_y , e $|d_y|$ (direita), calculadas em relação à orientação da grade (esquerda). Adaptado de (BAY et al., 2008). 20
- 2.14 Variação da detecção de pontos de interesse do SURF devido à mudanças no contraste. (a) Mesma imagem com contraste diferentes e quantidade de pontos detectados diferentes. 24, 75 e 78 respectivamente. (b) Imagens da mesma face sem pontos em comum. 21
- 2.15 Detecção de pontos pelo SURF em locais afastados mas dados como o mesmo ponto. 21

2.16 Classificação não-supervisionada. (a) Reconhecimento em uma classificação não-supervisionada onde S_i é uma distância vetorial, r é correspondente a mínima distância, e i indicando qual é a face mais próxima. (b) Verificação em uma classificação não-supervisionada onde V_p é o vetor de características da imagem face-de-busca, V_t é o vetor de características da imagem face-alvo, T é o limiar e r é a resposta do classificador que será “sim” caso o valor da distância $d(V_p, V_t)$ seja menor que o limiar ou “não” no caso contrário. 24

2.17 Classificação supervisionada. (a) Reconhecimento em uma classificação supervisionada onde V_j é uma concatenação dos vetore V_p , face-de-busca, e V_{t_j} , face-alvo, Ψ é uma função não linear, W são os pesos, b é o Baias, T é o limiar e C será a resposta da avaliação de V_j . V_p será reconhecido caso $\sum_j C \geq 1$ e a resposta do classificador será o j que a $t_j(V_j)$ seja igual 1. No caso de $\sum_j C = 0$ V_p não é conhecida. (b) Verificação em uma classificação supervisionada onde V é uma concatenação dos vetore V_p (características da imagem face-de-busca) e V_t (características da imagem face-alvo), ψ é uma função não linear, W são os pesos, b é o Baias, T é o limiar e $f(V)$ é a resposta do classificador que será “sim” caso o valor da função seja ≥ 0 ou “não” no caso contrário. 25

2.18 Exemplos de imagens normalizadas de uma mesma face do data set FERET. 27

2.19 Exemplos de imagens normalizadas de uma mesma face do data set IMM. 27

2.20 Exemplos de imagens normalizadas de uma mesma face do data set ORL. 28

2.21 Exemplos de imagens normalizadas de uma mesma face do data set Yale na primeira linha e do Extended Yale Face Database B na segunda linha. 29

2.22 Exemplos de imagens normalizadas de uma mesma face do data set PF83LFW. 30

3.1 Visão geral do COHAWES. (i) Aplicação dos filtros Haar wavelet, na vertical (d_y) e na horizontal (d_x), definida em uma região 20×20 para cada pixel e o cálculo de $\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|$ nas 16 sub-regiões. Vetorizando os valores obtidos em (i), obtemos o vetor v , de tamanho 64, em cada pixel; (ii) 25 grupos de vetores v são extraídos de acordo com a segmentação da imagem original; (iii) cálculo da matriz de correlação para cada grupo de vetores; (iv) vetorização do COHAWES a partir da concatenação dos valores acima da diagonal principal das matrizes de correlação (∇_S). . . . 34

3.2 Filtros Haar wavelet de tamanho 15×15 . (a) Filtros Haar wavelet para calcular as respostas na direção x (b) Filtros Haar wavelet para calcular as respostas na direção y . As partes escuras têm o peso -1 e as partes claras $+1$ 36

3.3	Os somatórios calculados em uma região representam o padrão de intensidade dos pixels. (a) Em uma região homogênea, todos os valores dos somatórios são relativamente baixos. (b) Em uma região que apresenta frequências na direção x , o valor de $\sum d_x $ é alto, mas todos os outros permanecem baixos. (c) A intensidade vai aumentando gradualmente na direção x , ambos os valores $\sum d_x$ e $\sum d_x $ são altos. Adaptado de (BAY et al., 2008).	36
3.4	Características extraídas dos somatórios das Haar wavelet. São produzidas 64 novas matrizes do mesmo tamanho da imagem.	37
3.5	O bloco de vetores v é dividido em grupos seguindo uma divisão da imagem original.	38
3.6	Divisões avaliadas durante o processo de construção do COHAWES.(a) Imagem inteira; (b) 2×2 ; (c) 4×4 ; (d) 5×5 ; (e) 7×7 ; (f) 8×8 ; (g) 10×10	38
3.7	Diagramas de dispersão. (a) Os valores de y tende a aumentar quando o valor de x cresce. (b) Os valores de y tende a decrescer quando o valor de x cresce. (c) Não apresenta correlação entre o valor de y o valor de x . (d) Existe uma correlação entre o valor de y o valor de x mas essa correlação não é linear. Adaptado de (LARSON; FARBER, 2010)	39
3.8	Em cor laranja, valores acima da diagonal principal da matriz de correlação que são considerados durante o processo de vetorização do COHAWES.	41
4.1	Acúmulo das distâncias entre os segmentos das imagens.	46
4.2	Utilização de um método de busca baseado em algoritmo genético na seleção do conjunto de características para representar as imagens de faces.	48
4.3	Etapas gerais do <i>framework</i> que permite a escolha do grupo das características, tipos de relação entre essas características, forma de divisão da imagem, classificador e data set.	51
4.4	Cenários de testes considerando 8 data sets, 7 descritores e 9 classificadores totalizando 504 cenários.	51
4.5	Comparação dos resultados do COHAWES com os outros descritores no data set FERET: FAFB, FADUP1 e FADUP2.	53
4.6	Comparação dos resultados do COHAWES com os outros descritores nos data sets IMM, ORL, PF83LFW, Yale e YaleB.	54
5.1	Diagramas de dispersão que possuem dispersões diferentes mas o valor da correlação é o mesmo.	61

LISTA DE TABELAS

2.1	Regras das imagem do data set IMM para cada pessoa.	28
4.1	Resumo dos data sets utilizados nas experimentações.	44
4.2	Precisão do reconhecimento de face para distância euclidiana, Manhattan e distância do cosseno com o vetor de 15 características v_g	47
4.3	Precisão do reconhecimento de face utilizando DCP. Dados da reprodução do trabalho (DING et al., 2016).	47
4.4	Percentuais de acerto de reconhecimento de identidade usando SVM como classificador e matriz de correlação do vetor v_g de 15 características como descritor.	48
4.5	Percentuais de acerto de reconhecimento de identidade utilizando o grupo de 5 característica selecionadas pelo algoritmo genético (v_{ag}).	49
4.6	Comparação do descritor utilizando matriz de correlação dos somatórios de gradientes de filtros do tipo Haar wavelet (COHAWES) e divisão 5×5 com correlação do vetor com 15 característica (v_g) e divisão 5×5 , correlação do vetor com as característica selecionadas pelo algoritmo genético (v_{ga}) e divisão 5×5 e DCP (DING et al., 2016). Classificador implementação SVM de (ORTIZ; BECKER, 2014).	50
4.7	Precisão de reconhecimento facial de todos os descritores e todo os classificadores no FERET.	53
4.8	Precisão de reconhecimento facial de todos os descritores e todos os classificadores nos data sets IMM, ORL, PF83LFW, Yale e Yale <i>extended</i> B.	55
4.9	Média dos percentuais de acerto para cada descritor.	56

LISTA DE ALGORITMOS

3.1 Pseudocódigo do COHAWES	35
---------------------------------------	----

LISTA DE ABREVIATURAS E SIGLAS

COHAWES	Correlation of Haar wavelet Sums
DCP	Dual-Cross Patterns
FERET	Facial Recognition Technology
GPSR	Gradient Projection for Sparse Reconstruction
HOG	Histogram of Oriented Gradients
KNN-SRC	K Nearest Neighbor based Local Sparse Representation Classifier
LASRC	Linearly Approximated Sparse Representation-based Classification
LBP	Local Binary Patterns
LLC	Locally-constrained Linear Coding
LRC	Linear Regression Classification
LSHOG	Locality Sensitive Histograms of Oriented Gradients
LTP	Local Ternary Patterns
NN	Nearest Neighbor
SRC	Sparse Representation-based Classification
SURF	Speed Up Robust Features
SVM	Support-Vector Machines



INTRODUÇÃO

1.1	Motivação	4
1.2	Objetivos	4
1.3	Contribuições	5
1.4	Mapa dos capítulos	5

Informações biométricas têm sido utilizadas frequentemente para a tarefa de reconhecimento automático de pessoas. Entretanto, as técnicas mais comuns necessitam de uma ação do indivíduo para fornecer essas informações, como no caso de colocar o dedo em um lugar específico e ter a impressão digital lida, ou se aproximar e olhar fixamente para um ponto e ter a íris reconhecida.

Para realizar a tarefa de reconhecimento facial, de forma automática, basta a pessoa ter a imagem da face capturada por uma câmera para que o processo de reconhecimento possa ser executado, por exemplo. Essa captura pode ser feita ainda sem participação ativa do indivíduo, como nos casos dos sistemas de vigilância onde o usuário nem tem conhecimento de ter sido identificado. Esse reconhecimento de forma passiva da face de um usuário pode ser feito em ambientes como aeroportos, terminais rodoviários, estádios, dentre outros, com o objetivo de prover uma maior segurança nesses locais. O reconhecimento facial pode ser utilizado também em situações nas quais o acesso físico, em determinados ambientes, é restrito, como por exemplo em sala-cofre, laboratórios, departamentos que tenham informações críticas, dentre outros. Atualmente, mesmo para desbloquear o acesso a dispositivos móveis, já existem sistemas de reconhecimento facial que identifica quem possui acesso aquele aparelho.

Quando o reconhecimento facial ocorre com o objetivo de autorizar o acesso de alguém a algum local, normalmente a câmera, que será utilizada para realizar a captura da imagem da face, poderá ficar em um local mais próximo do indivíduo, permitindo ter um controle melhor sobre as condições de iluminação e pose de captura da face. Isso contribui para aumentar a precisão do sistema de reconhecimento facial, já que alguns

dos desafios desses sistemas poderão ser minimizados, tais como iluminação inadequada, face rotacionada, baixa qualidade da imagem, apenas para citar alguns.

Com o aumento do poder de processamento e da capacidade de armazenamento dos equipamentos computacionais e das câmeras digitais, juntamente com a diminuição dos custos desses aparatos, tornou-se muito mais barato adquirir os equipamentos necessários para construir sistemas de vídeo-vigilância. Esses sistemas podem ser encontrados em pequenas empresas e até mesmo em ambientes domésticos. Muitos equipamentos para interfone do tipo “porteiro eletrônico” possuem câmeras e transmissão das imagens para que alguém possa identificar a pessoa que deseja entrar. Algumas redes sociais e softwares de gerenciamento e armazenamento de imagens realizam o reconhecimento facial para o agrupamento das imagens. Normalmente em situações que envolvem sistemas de vídeo-vigilância, a câmera fica em um local mais afastado do local a ser monitorado para, além de capturar a imagem das faces, conseguir ter uma percepção mais geral do ambiente. Nesses casos, o reconhecimento de forma automática é mais desafiador já que as imagens da face podem não estar com a iluminação e na posição que o sistema de reconhecimento necessite para realizar a identificação.

O processo de reconhecimento facial é um problema de reconhecimento de padrões visuais no qual a face é representada como um objeto sujeito a iluminação variada, poses diversas, expressões faciais e outras situações, e esta precisa ser identificada com base em imagens previamente adquiridas. Um sistema de reconhecimento de face consiste geralmente em quatro módulos que são: detecção da face, normalização, extração das características e classificação (LI; JAIN, 2011). Por se tratar de um problema que envolve várias etapas, algumas pesquisas se focaram em módulos específicos do processo de reconhecimento facial. Como o caso de Osuna, Freund e Giroit (1997), que utilizaram uma *support-vector machines* (SVM) para separar imagens de face de imagens que não eram de face, e também Viola e Jones (2001), que desenvolveram um detector de objeto que é capaz de processar imagens de forma extremamente rápida e alcançar altas taxas de detecção utilizaram um algoritmo de aprendizagem baseado em AdaBoost. Para a etapa da normalização, existem trabalhos específicos como o de Li et al. (2006), onde os autores fazem uso do detector de Viola e Jones (2001), como um detector de face, e depois localiza a posição das pupilas usando uma relação geométrica entre o rosto e os olhos; depois desta etapa, o método proposto normaliza a orientação da face, realizando uma rotação da imagem para que os olhos fiquem alinhados com o eixo horizontal. O detector de Viola e Jones (2001) realiza também uma normalização da escala da imagem, redimensionando a imagem da face para que fique sempre com o mesmo tamanho e ajustando o brilho para que todas as imagens tenham o mesmo valor médio dos pixels e mesma variância de valores.

A extração das características é uma das etapas mais importantes para problemas de detecção e classificação. Características devem ser discriminativas, robustas e rápidas de se calcular. Os valores dos pixels é a escolha mais simples para serem utilizados como características, no entanto, esses valores sofrem uma influência muito grande das variações na iluminação, mudanças de pose e rotações. Os pixels crus também sofrem do problema de grande dimensionalidade dos dados para serem processados pelos classificadores (TUZEL; PORIKLI; MEER, 2006) de forma rápida. Diversas estratégias foram

aplicadas para métodos de seleção e extração de características, buscando aumentar a robustez das características extraídas das imagens da face. Dentre essas estratégias, algumas das mais utilizadas são: Gabor wavelets (LIU; WECHSLER, 2002), *local binary patterns* (LBP) (AHONEN; HADID; PIETIKAINEN, 2006), *histogram of oriented gradients* (HOG) (DALAL; TRIGGS, 2005) e mais recentemente o *dual-cross patterns* (DCP) (DING et al., 2016). O *speed up robust features* (SURF) é um detector e descritor de ponto de interesse e foi apresentado por Bay, Tuytelaars e Gool (2006) como apresentando um certo grau de invariância em escalamento e rotação, com vantagens na velocidade a qual pode ser calculado e que se aproxima, ou mesmo supera, os esquemas propostos anteriormente em relação à repetibilidade, distinção e robustez (BAY; TUYTELAARS; GOOL, 2006). O SURF também foi utilizado para reconhecimento facial em Du, Su e Cai (2009) e Anand e Shah (2016).

Métodos de aprendizagem de máquinas são aplicados para realizar a separação entre as faces de um mesmo indivíduo das faces dos demais. Esses métodos são aplicados para o processo de reconhecimento facial e alguns dos classificadores mais utilizados em sistemas de reconhecimento facial são: *nearest neighbor* (NN) (COVER; HART, 1967), *locally-constrained linear coding* (LLC) (WANG et al., 2010), *k-nearest neighbor com sparse representation classifier* (KNN-SRC) (ZHANG; YANG, 2010), *linear regression classification* (LRC) (NASEEM; TOGNERI; BENNAMOUN, 2010), minimização ortonormal ℓ_2 (MO-L2) (SHI et al., 2011), *support-vector machines* (SVM) (HEISELE; HO; POGGIO, 2001), *linearly approximated sparse representation-based classification* (LASRC) (ORTIZ; BECKER, 2014), *sparse representation classifier with gradient projection for sparse reconstruction* (SRC-GPSR) (FIGUEIREDO; NOWAK; WRIGHT, 2007), e *sparse representation classifier with homotopy* (SRC-Homotopy) (MALIOUTOV; CETIN; WILLSKY, 2007).

Nosso trabalho está inserido no módulo de extração de características, com a proposta de um novo descritor para representar a face, o *correlation of haar wavelet sums* (COHAWES). Este descritor calcula o somatório de gradientes da aplicação de filtros tipo Haar wavelet em uma vizinhança de 20×20 , de forma densa, criando um vetor com tamanho 64 para cada pixel da imagem. Em seguida, a imagem é dividida em regiões e, em cada região, é calculada a correlação entre as 64 dimensões do vetor. Todos os descritores e classificadores citados serão tratados mais detalhadamente no Capítulo 2 e foram utilizados para comparação com o descritor proposto, no processo de reconhecimento facial.

Para a etapa de reconhecimento facial propriamente dito, foram utilizados alguns data sets de face para avaliar o COHAWES. Para tanto, os data sets utilizados foram: Facial Recognition Technology (FERET) (PHILLIPS et al., 1997), The ORL Database of Faces (CAMBRIDGE, 2002), The IMM Face Database (NORDSTRØM et al., 2004), o Yale e o Extended Yale Face Database B (GEORGHIADES; BELHUMEUR; KRIEGMAN, 2001) e uma junção dos data sets LFW (HUANG et al., 2007) e PubFig83 (PINTO et al., 2011), proposta por Becker e Ortiz (2013).

Para auxiliar os experimentos, e realizar as comparações do COHAWES com os outros descritores nos diversos classificadores, foi desenvolvido um *framework* que permite a parametrização de vários passos do sistema de reconhecimento facial. Com este *framework*,

foram parametrizados os descritores que seriam utilizados no processo, os classificadores, os data sets e as possíveis variações do COHAWES com as diferentes formas de dividir a imagem em regiões, conjuntos de características que seria considerado no cálculo da correlação e a possibilidade de substituir a correlação pela covariância.

1.1 MOTIVAÇÃO

Apesar dos avanços trazidos pelas técnicas recentes, e dos avanços tecnológicos dos equipamentos computacionais, o reconhecimento facial ainda é um problema a ser resolvido. Algumas abordagens apresentam resultados promissores em situações específicas, tais como: em iluminação uniforme, expressão facial neutra, posicionamento frontal da face para a câmera, sem inclinações e face completamente visível.

Avanços na área de reconhecimento facial juntamente com os baixos custos das câmeras digitais, a portabilidade e acessibilidade a esses equipamentos, permitem a aplicação do reconhecimento automático da face nos mais diversos contextos, inclusive, possibilitando uma interação mais natural com os sistemas computacionais.

Zhao et al. (2003) diz que as duas questões cruciais que precisam ser resolvidas em sistemas de reconhecimento facial são: a representação da face e a estratégia do classificador. O objetivo da representação da face é ter um descritor que consiga discriminar variações entre indivíduos diferentes, mas que não seja sensível a variações nas imagens de um mesmo indivíduos. Sem um descritor que codifique as variações entre indivíduos, os classificadores não conseguirão separar as classes e o reconhecimento facial não será feito.

De acordo com Raschka (2014), a distribuição normal padrão dos valores dos descritores é um requisito geral para muitos algoritmos de aprendizagem de máquinas. Sem essa padronização, certos valores do vetor de características em diferentes escalas, podem atualizar mais rapidamente do que outros, o que influenciaria no resultado final da classificação. Juntando isso ao que foi proposto por Tuzel, Porikli e Meer, com a utilização das matrizes de covariância, este trabalho apresenta uma proposta de descritor que utiliza a matriz de correlação e tem como resultado um vetor com os seus valores em uma distribuição normal, com média 0 e desvio padrão 1. Outro aspecto considerado é que o SURF apresenta um descritor robusto e com baixos custo computacional, mas a etapa de detecção dos pontos de interesse não possui uma relação com os pontos fiduciais. Isso faz com que, por algumas vezes, os pontos detectados pelo SURF não sejam discriminantes o suficiente para o classificador conseguir reconhecer a face. Estas foram as principais motivações para a concepção do COHAWES.

1.2 OBJETIVOS

O presente trabalho propõe um novo descritor para ser utilizado em sistemas de reconhecimento facial e com esse descritor conseguir uma representação da face que evidencie nuances em imagens de indivíduos diferentes, mas que não sofra grandes modificações em imagens de um mesmo indivíduo.

Como objetivos específicos, os seguintes itens devem ser mencionados:

- Desenvolver um descritor cujos valores estejam naturalmente dentro da distribuição normal padrão;
- Considerar toda a imagem para o cálculo do descritor;
- Desenvolver um *framework* para realização de testes de reconhecimento facial a fim de se obter uma visão mais realista do desempenho do descritor proposto.

1.3 CONTRIBUIÇÕES

Como contribuição, este trabalho propõe o novo descritor COHAWES que sintetiza informações de textura da imagem considerando a vizinhança de cada pixel. Essa síntese é feita juntamente com a divisão da imagem em regiões para que as informações de disposição espacial sejam consideradas no descritor. Além disso, os valores do COHAWES estão dentro da distribuição normal padrão. Este trabalho apresenta também um *framework*, com o código-fonte disponível, para facilitar a avaliação do desempenho do descritor proposto em tarefas de reconhecimento facial, parametrizando conjuntos de características, tipo de relação entre as características (correlação e covariância), formas de segmentação da imagem, classificadores e data sets. Por fim, outra contribuição desse trabalho é a realização de uma avaliação do poder de discriminação de descritores que utilizam covariância e correlação de características. Essa avaliação foi feita utilizando os mesmos conjuntos de dados e os mesmos classificadores.

1.4 MAPA DOS CAPÍTULOS

Este trabalho segue a seguinte estrutura de capítulos:

- O **Capítulo 2** faz uma apresentação da fundamentação teórica sobre o reconhecimento facial e principais descritores utilizados e alguns dos data sets mais comuns utilizados para avaliar sistemas de reconhecimento facial. Ao final, a relação dos trabalhos do estado-da-arte é analisada com face ao descritor proposto.
- O **Capítulo 3** apresenta detalhes técnico sobre o COHAWES, a sua estrutura e os seus parâmetros.
- O **Capítulo 4** descreve a metodologia usada para validação dos testes no trabalho proposto, apresenta os resultados obtidos e compara o descritor proposto com outros descritores aplicando o *framework* proposto.
- O **Capítulo 5** apresenta as considerações finais do trabalho, as principais contribuições, limitações e sugestões para trabalhos futuros.

SISTEMAS DE RECONHECIMENTO FACIAL

2.1	Definições gerais	8
2.2	Extratores de características	10
2.2.1	Gabor wavelets	11
2.2.2	<i>Local binary patterns</i>	14
2.2.3	<i>Histogram of oriented gradients</i>	15
2.2.4	<i>Dual-cross patterns</i>	16
2.2.5	SURF	17
2.2.6	Descritor de covariância	22
2.3	Classificadores	23
2.3.1	Classificação não-supervisionada	23
2.3.2	Classificação supervisionada	24
2.4	Data sets	26
2.4.1	Facial recognition technology	27
2.4.2	The IMM face database	28
2.4.3	The ORL database of faces	29
2.4.4	Yale data sets	29
2.4.5	PubFig83+LFW (PF83LFW)	30
2.5	Relação com o trabalho proposto	30

Reconhecer faces é uma tarefa que as pessoas realizam diariamente. Entretanto, reproduzir automaticamente tal tarefa ainda tem sido objeto de vários estudos. Essa habilidade humana de identificar pessoas a partir da face ainda não foi completamente compreendida, tornando mais difícil a sua reprodução de maneira robusta.

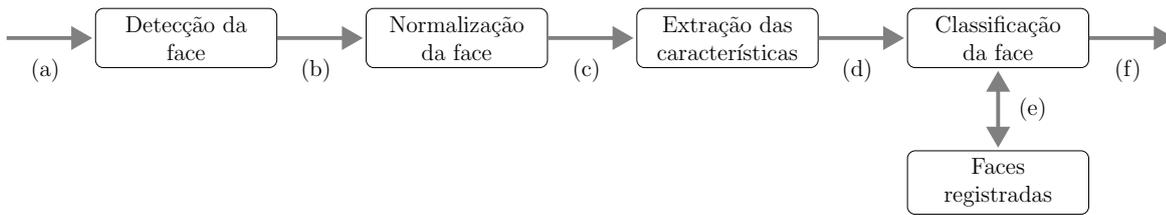


Figura 2.1: Fluxo geral de um sistema de reconhecimento facial. (a) Imagem a ser identificada, (b) face detectada com pontos fiduciais, (c) imagem unicamente da face, (d) representação da face a partir de um vetor de características, (e) busca da face no banco de faces e (f) identificação da face propriamente dita.

Os primeiros estudos registrados no âmbito do reconhecimento de face datam da década de 1960, (BLEDSOE, 1966) e (GOLDSTEIN; HARMON; LESK, 1971). Na época, o sistema de reconhecimento era semiautomatizado e exigia que um administrador localizasse as características (como olhos, ouvidos, nariz e boca) nas fotografias antes de calcular distâncias e proporções para um ponto de referência comum. Eram utilizadas 21 características, como a cor do cabelo e a espessura dos lábios, para realizar a comparação. O primeiro sistema automatizado para reconhecimento facial é creditado a Kanade, em seu trabalho de tese de doutorado (KANADE, 1973). Kanade usava as distâncias entre as características locais (olhos, nariz e boca) para calcular semelhanças entre rostos. Essa estratégia é baseada na detecção de características locais, utilização de propriedades geométricas e a análise e comparação dessas propriedades geométricas para classificação da imagem.

No final da década de 1980 (SIROVICH; KIRBY, 1987), e início da década de 1990 (KIRBY; SIROVICH, 1990), Kirby e Sirovich propuseram uma forma automática de extração de características da imagem da face aplicando uma técnica de redução dimensional chamada de *Karhunen-Loève expansion*, *Principal Component Analysis* ou *hotelling transform*. As características extraídas eram resultantes das autofunções da matriz de covariância do conjunto de faces; tais características passaram a ser chamadas de *Eigenpictures*.

2.1 DEFINIÇÕES GERAIS

O reconhecimento de face se apresenta como um problema de reconhecimento de padrões visuais, onde a face está sujeita a variações de iluminação, pose, expressão, entre outros. Esse objeto-face na imagem precisa ser identificado com base em imagens previamente adquiridas (LI; JAIN, 2011). Um sistema de reconhecimento de face consiste geralmente em quatro módulos: detecção da face, normalização, extração de características e classificação. A Figura 2.1 ilustra o fluxo geral usualmente adotado em um sistema de reconhecimento facial.

O termo “reconhecimento de face” pode ser aplicado em dois cenários diferentes: “reconhecimento” (ou “identificação”) e “verificação” (ou “autenticação”). Em ambos os cenários, um conjunto de faces com identidades conhecidas é inicialmente cadastrado



Figura 2.2: (a) Estratégia de reconhecimento: o sistema busca a identidade de uma face-de-busca a partir de uma galeria com faces-alvo. (b) Estratégia de verificação: uma face-alvo reivindica uma identidade ao sistema, o qual responde que sim ou não, caso a face-alvo tenha acesso ou não ao referido sistema, respectivamente.

numa galeria de imagens (do Inglês, *gallery*). As imagens da galeria são chamadas de faces-alvo (do Inglês, *target*) (BOWYER; CHANG; FLYNN, 2006). Após a inserção das faces-alvo na galeria, outras imagens serão utilizadas para fazer buscas no sistema de reconhecimento facial com o objetivo de encontrar a identidade de uma determinada pessoa; tais imagens são conhecidas como face-de-busca (do Inglês, *probe*).

Em um cenário de reconhecimento, a correspondência da face é de um-para-muitos, no sentido de que uma face-de-busca é comparada com todas as faces-alvo da galeria, a fim de encontrar a melhor combinação acima de um limiar (do Inglês, *threshold*) (BOWYER; CHANG; FLYNN, 2006). A Figura 2.2(a) ilustra este cenário de reconhecimento. Um exemplo prático de reconhecimento facial pode ser encontrado em um sistema de registro de entrada e saída em algum ambiente controlado, onde pelo vídeo-monitoramento, um sistema de reconhecimento facial registra quem acessou o ambiente e em que momento.

Em um cenário de verificação, a correspondência é de um-para-um, no sentido de que a face-de-busca é comparada com apenas uma face-alvo na galeria, com o objetivo de se conceder ou não acesso ao sistema. A verificação é realizada quando a comparação entre a face-de-busca e o face-alvo for acima de um limiar (BOWYER; CHANG; FLYNN, 2006). A Figura 2.2(b) ilustra o cenário de verificação. Um exemplo de verificação de identidade é um sistema de controle de acesso em um dispositivo móvel utilizando a face do usuário.

Tanto em cenários de reconhecimento, quanto em cenários de verificação, duas questões cruciais precisam ser resolvidas: a representação da face e a estratégia do classificador. O objetivo da representação da face é discriminar variações entre indivíduos diferentes, mantendo a mesma representação para faces do mesmo indivíduo. Sem um descritor (seja este treinável ou definido previamente a partir de uma função) que represente as variações entre indivíduos, os classificadores não conseguirão separar adequadamente as classes (ZHAO et al., 2003). A representação de objetos em imagens é realizada a partir de um extrator de características, o qual é obtido a partir de um treinamento ou previamente definido. Na Seção 2.2 alguns extratos pré-definidos utilizados neste trabalho são apresentados. Foge ao escopo deste trabalho avaliar os extratores treináveis (do tipo aqueles obtidos por estratégias *deep learning*). Quanto a estratégia de classificação, esta diz respeito ao caráter supervisionado ou não-supervisionado da classificação das características de imagens extraídas, e será discutida na Seção 2.3.

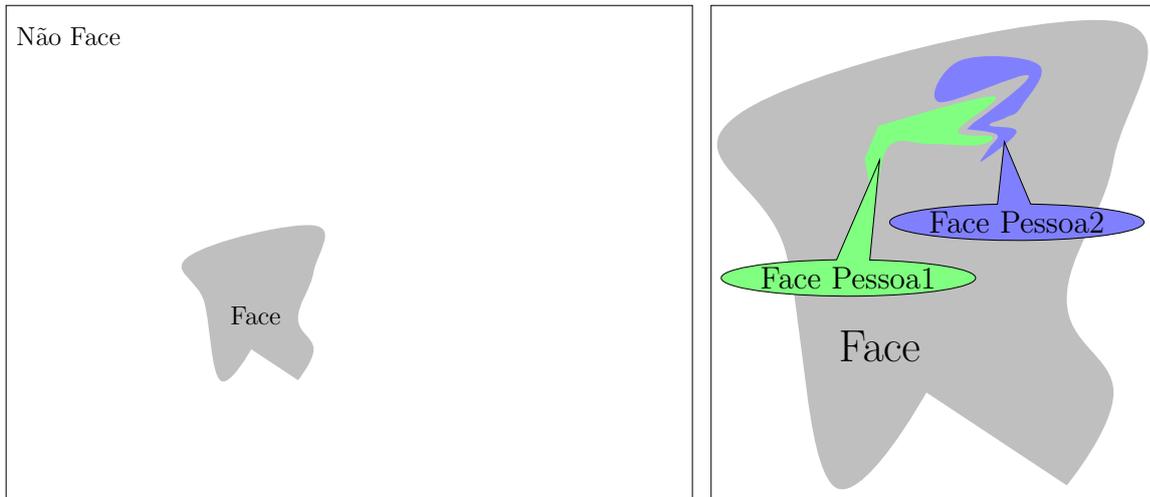


Figura 2.3: Espaço de representação de objetos não-face, face e dentro de face a separação entre faces de pessoas diferentes. Adaptado de (LI; JAIN, 2011).

2.2 EXTRATORES DE CARACTERÍSTICAS

A eficácia de um sistema de reconhecimento de face depende fortemente das características que são extraídas para representar o padrão da face e dos métodos de classificação utilizados para distinguir as faces de diferentes pessoas. Essas representações da face devem compor um espaço onde faces de uma mesma pessoa pertençam ao mesmo subespaço de características, enquanto faces de pessoas diferentes pertencem a subespaços distintos. A Figura 2.3 ilustra o espaço de representação de objetos não-face, na esquerda e na direita, dentro da representação de objetos face, o destaque para a representação da face da Pessoa 1, em verde, e Pessoa 2 em azul.

O objetivo perseguido em um extrator de características é usualmente fazer com que os subespaços de características de faces de pessoas diferentes sejam tão linearmente separáveis quanto possível, para que os métodos de classificação possam separar os subespaços. Uma estratégia que auxilia alcançar tal objetivo é fazer um processamento da imagem antes da extração das características e nesse pré-processamento a imagem da face deve ser normalizada quanto aos aspectos geométricos (como inclinação da face) e fotométricos (como iluminação). Apesar dessa normalização, as características extraídas devem possuir também alguma invariância a algumas transformações geométricas e fotométricas (LI; JAIN, 2011).

Após a normalização, a etapa de extração das características pode considerar a imagem toda ou apenas regiões específicas. Seguindo essa ideia, Zhao et al. (2003) definiram três categorias de abordagem para extração de características: Holísticas, estruturais e híbridas. Na Figura 2.4, são ilustrados alguns exemplos das abordagens holísticas e estruturais. A categoria híbrida utiliza a abordagem de características globais da categoria holística conjuntamente com características locais da categoria estrutural.

Baseado nos trabalhos de Ortiz e Becker (2014) e Ding et al. (2016), utilizamos os

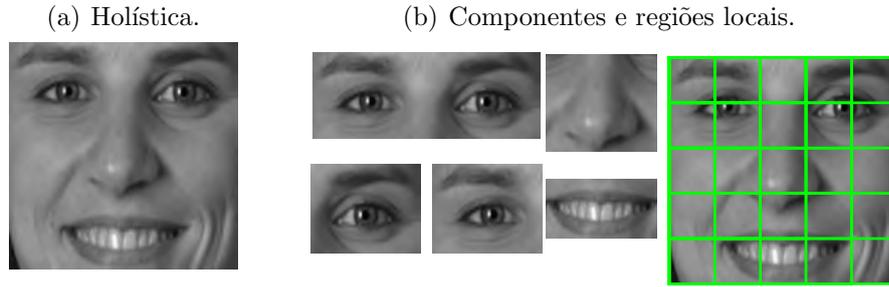


Figura 2.4: Exemplos de abordagens para extração de características de uma imagem de face: (a) Apresenta a abordagem holística onde toda a face é considerada. (b) Apresenta a abordagem estrutural onde componentes da face ou regiões da imagem são consideradas

extratores de características do tipo: *Gabor wavelets* (LIU; WECHSLER, 2002), *local binary patterns* (LBP) (AHONEN; HADID; PIETIKAINEN, 2006), *histogram of oriented gradients* (HOG) (DALAL; TRIGGS, 2005), *dual-cross patterns* (DCP) Ding et al. (2016), para comparação com o extrator de característica proposto em nosso trabalho. Além disso, o extrator speedup robust features (SURF) (BAY; TUYTELAARS; GOOL, 2006) é apresentado por também utilizar os somatórios de gradientes de filtros do tipo Haar wavelet e o descritor de covariância (TUZEL; PORIKLI; MEER, 2006) por ter utilizado uma medida do grau de interdependência entre características.

2.2.1 Gabor wavelets

Gabor wavelets são filtros lineares usados para a análise de textura com o objetivo de encontrar características na imagem em frequências e em direções específicas. O filtro 2D de Gabor possui uma componente Gaussiana que possibilita alterações na escala do filtro, e uma componente senoidal, que permite definir uma frequência de interesse para análise a imagem. A modulação de uma Gaussiana por uma senoidal é dada por:

$$\varphi_{\Pi(f,\theta,\gamma,\eta)}(x,y) = \frac{f^2}{\pi\gamma\eta} \exp(-(\alpha^2 x'^2 + \beta^2 y'^2)) \exp(j2\pi f x'), \quad (2.1)$$

$$x' = x \cos \theta + y \sin \theta,$$

$$y' = -x \sin \theta + y \cos \theta,$$

onde f é a frequência da modulação senoidal, θ é a rotação da gaussiana, α e β são os valores relacionados com a nitidez da gaussiana em x e y e para manter essa nitidez constante são usados $\gamma = f/\alpha$ e $\eta = f/\beta$. Configurando os parâmetros para determinar a escala, a rotação e a frequência, é definido um núcleo de Gabor. Geralmente é definida uma família de núcleos de Gabor e esses núcleos são convolucionados com o sinal original, resultando em um chamado espaço de Gabor.

A utilização de Gabor wavelets para reconhecimento de faces teve seu início com o trabalho de Lades et al. (1993), onde a face é representada por um grafo e cada vértice desse grafo armazena um conjunto de características baseadas em Gabor wavelets. Para

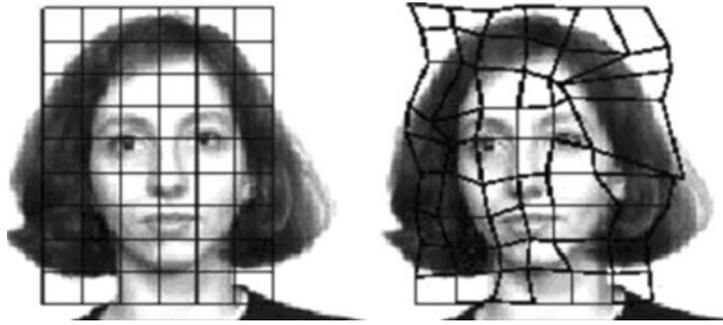


Figura 2.5: Representação dos grafos que descrevem duas imagens de face. Representação de uma face através de um grafo. Quando essas distorções estão dentro do valor de aceite as imagens são consideradas da mesma face. Figura extraída de (LADES et al., 1993).

o reconhecimento da face, os grafos que representam as imagens eram deformados até os nós combinarem. Quando os nós estavam ajustados e a distorção do grafo estava dentro de um valor de aceite era considerado que as duas imagens eram da mesma face. Uma ilustração desse processo pode ser vista na Fig. 2.5. Uma abordagem semelhante pode ser encontrada no trabalho de Wiskott et al. (1997), onde uma estrutura chamada de *elastic bunch graph matching* foi definida para representar a face. Em cada ponto dessa estrutura, são extraídas características locais baseadas em uma família de Gabor wavelets.

Gabor wavelets também foram aplicadas em identificações de face baseadas em abordagens holísticas para o reconhecimento facial. Liu e Wechsler (2002) utilizaram cinco escalas diferentes e oito orientações, gerando um banco de quarenta núcleos de Gabor. Liu e Wechsler realizaram a convolução da imagem original da face com cada um dos núcleos gerando assim quarenta novas imagens. Cada imagem nova destaca o local da imagem original que tenha uma textura semelhante ao do núcleo de Gabor que a imagem fez a convolução. Uma amostragem de resolução mais baixa é feita para cada uma dessas novas 40 imagens e a junção dessas amostragens criam a nova representação da imagem da face. Os núcleos utilizados por Liu e Wechsler, bem como o resultado da convolução e a nova representação da face podem ser vistos na Fig. 2.6.

Ayinde e Yang (2002) utilizaram três combinações de escalas e orientações para definir os núcleos de gabor. Em um núcleo é utilizado um filtro de cosseno com tamanho 5×5 e orientações em 0° , um segundo núcleo é com um filtro de seno com tamanho 5×5 e orientações em 90° e um terceiro núcleo com filtro de cosseno com tamanho 9×9 e orientações em 0° .

Os filtros de Gabor são particularmente eficientes para extrair recursos de imagens de rosto devido à sua capacidade de adquirir propriedades visuais significativas, como a localização espacial, a frequência e a orientação, fazendo com que esses filtros sejam frequentemente usados no reconhecimento facial. Seguindo a estratégia de grafos com nós de Lades et al. (1993), em Cament et al. (2015) fizeram um grafo com 4172 nós para realizar o reconhecimento da face. Esse processo gerou um aumento significativo no tempo de processamento.

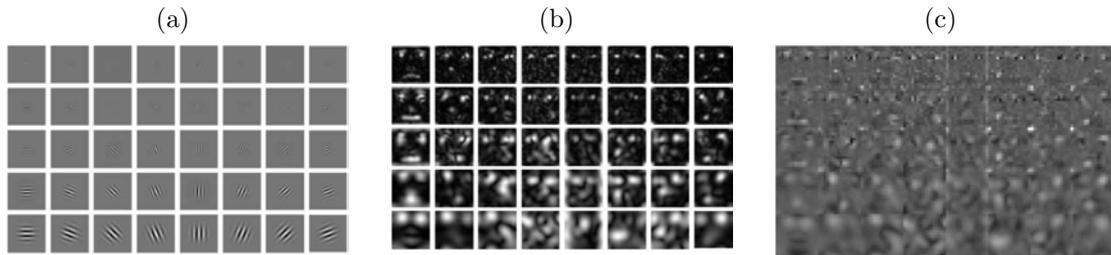


Figura 2.6: Aplicação de núcleos de Gabor de forma holística na face. (a) 40 núcleos de Gabor obtidos por uma variação de 5 escalas diferentes e 8 orientações. (b) Convolução da imagem da face com cada um dos núcleos de Gabor definidos. (c) Junção da amostragem em baixa resolução das imagens geradas pela convolução da imagem da face com os núcleos de Gabor criando características para representação da face. Figura extraída de (LIU; WECHSLER, 2002)

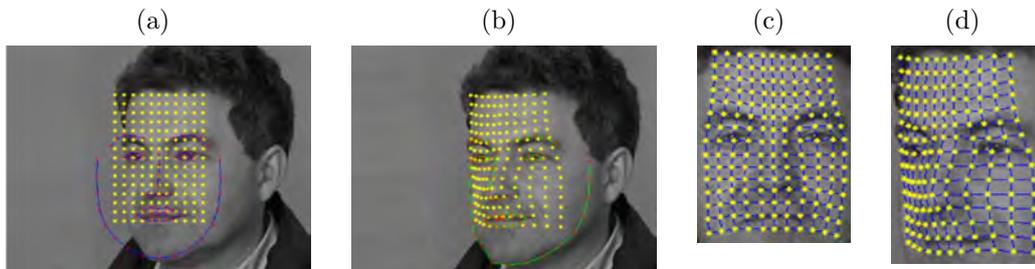


Figura 2.7: Exemplo de deformação das malha utilizadas para calcular os pontos de convoluções com os filtros de Gabor. (a) Exibição da malha inicial (em amarelo) e forma média da face (em azul). (b) Forma padrão da face ajustada (verde) com a da imagem e malha sofrendo as mesmas deformações (amarelo) fazendo com que a posição dos pontos de convoluções com os filtros de Gabor tenham a mesma posição relativa às características de uma face em diferentes imagens. As figuras (c) e (d) mostram a posição dos pontos de convoluções com os filtros de Gabor em face frontal e rotacionada, respectivamente. Figura extraída de (CAMENT et al., 2015).

No trabalho de Cament et al. (2015) apresentam uma estratégia de extração das características da face baseada em uma malha de pontos, na qual em cada ponto serão extraídas as convoluções com os filtros de Gabor. Essa malha se molda à face e se ajusta para conseguir identificar variações da pose. Primeiramente, é realizada uma detecção de pontos na face para se obter um alinhamento com uma forma de contorno de face padrão utilizada pelo método proposto. Com esse alinhamento feito, a malha é ajustada para corresponder às mesmas deformações que o formato de face padrão sofreu para se ajustar aos pontos detectados. Esse alinhamento pode ver visto da Fig. 2.7.

Após o alinhamento são definidas 5 malhas com escalas diferentes. Em cada ponto das malhas são realizadas convoluções da imagem com filtros de 8 orientações diferentes.

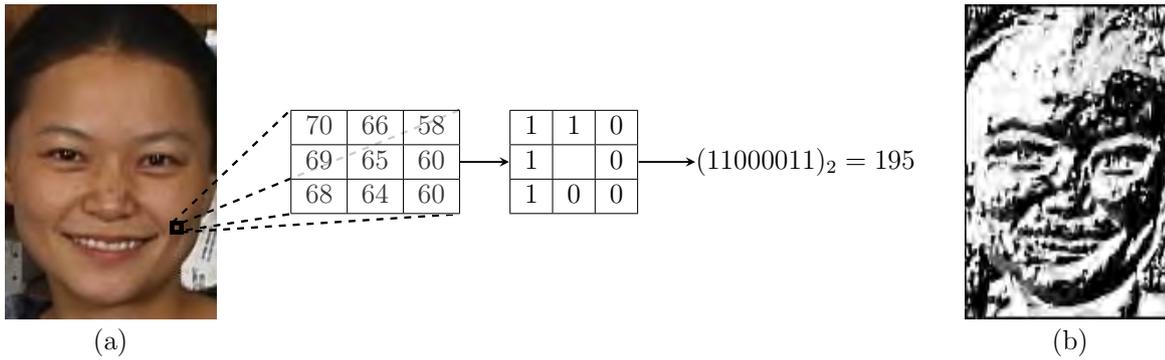


Figura 2.8: Cálculo do Local Binary Patterns. (a) Imagem Original. Ao lado da imagem original é mostrado um exemplo de uma vizinhança 3×3 . Todos os pixels que tiveram um valor maior, ou igual, ao pixel central receberá o valor 1 e os que tiverem valores menores receberá 0. A sequência de 0s e 1s é convertida para base 10 (b) Imagem após cada pixel ser substituído pelo LBP.

O conjunto das convoluções do ponto da malhas é chamado de *jet* e cada *jet* tem um vetor com tamanho 8. Os *jets* também são extraídos de uma normalização da imagem gerando um novo conjunto de características. Os *jet* são comparados com os extraídos previamente das imagens da base de dados para realizar o reconhecimento.

2.2.2 Local binary patterns

Ojala, Pietikäinen e Harwood (1996) apresentaram um novo descritor de texturas com base na subtração entre um pixel central de uma região de 3×3 e seus vizinhos. A subtração, $s(x)$, entre o pixel central x e seus vizinhos, é dada por:

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{caso contrário} \end{cases} \quad (2.2)$$

Depois de realizar a soma, todos os valores dos pixels vizinhos são concatenados em uma palavra binária e esse valor é convertido para a base decimal, conforme representado na Fig. 2.8.

O LBP é muito utilizado em sistemas de reconhecimento de faces, como nos trabalhos de Ahonen et al. (2004), Ahonen, Hadid e Pietikainen (2006), que dividem a imagem da face em regiões menores e calculam o histograma do LBP em cada região. Os histogramas são concatenados e formam o vetor de características que representa a face. Algumas variações em cima do método LBP surgiram, como nos trabalhos de Tan et al. (2007), Nagaraju, Srinu e Rao (2011), onde um novo critério foi acrescentado na Eq. 2.2, fazendo com que os valores possíveis passassem a ser -1, 0 e 1. Isto gerou um novo método de representação denominado *local ternary patterns* (LTP).

Zhang et al. (2005) primeiro decompõe a imagem da face a partir da aplicação de filtros de Gabor gerando imagens de magnitudes de Gabor, em seguida calcula o LBP

sobre essas magnitudes. Cada LBP resultante é dividido em regiões e nessas regiões são calculados os histogramas. Com a junção dos histogramas, um modelo de representação da face é construído. Uma outra combinação do LBP com o Gabor foi no trabalho de Tan e Triggs (2007), onde calcularam o LBP e o Gabor de forma independente e depois combinaram os valores dos dois extratores.

2.2.3 Histogram of oriented gradients

O *histogram of oriented gradients* (HOG) é um extrator que tem como característica codificar a distribuição das magnitudes dos gradientes a partir das orientações desses gradientes. De acordo com Dalal e Triggs (2005), para realizar o cálculo do HOG devem ser seguidos os seguintes passos: (i) calcular a magnitude e a orientação dos gradientes da imagem; (ii) realizar uma divisão da imagem em células e blocos; (iii) calcular o histograma de gradientes por células, agrupados em blocos (iv) concatenar esses histogramas em um vetor que será então o descritor HOG.

Os gradientes são calculados com a aplicação de um filtro de detecção de bordas de Sobel em 1D. Esse filtro é definido como $\begin{bmatrix} 1 & 0 & 1 \end{bmatrix}$ e é aplicado na horizontal (em x) e, sua transposta na vertical (em y) gerando os valores de gradientes g_x e g_y para toda a imagem. Com esses dois gradientes é calculado a magnitude de cada pixel da seguinte forma:

$$g = \sqrt{g_x^2 + g_y^2}, \quad (2.3)$$

onde g é o valor da magnitude do gradiente em um dado pixel. Além da magnitude do gradiente, g_x e g_y são utilizados para determinar a direção desse gradiente da seguinte forma:

$$\theta = \arctan \frac{g_y}{g_x}. \quad (2.4)$$

Após o cálculo das magnitudes e das direções dos gradientes o próximo passo é dividir a imagem em células e em blocos. As células são subdivisões da imagem em regiões com o mesmo tamanho e os blocos são conjuntos sobrepostos de 2×2 células.

No terceiro passo, as direções dos gradientes são normalizadas no intervalo de $0^\circ - 180^\circ$ e essa faixa de representação da direção é dividida em 9 *bins*¹. As magnitudes dos gradientes de cada bloco devem ser somadas em histogramas com essas 9 divisões. Ao final, devem ser concatenados todos os histogramas para a montagem do HOG. Esse processo pode ser melhor compreendido na Fig. 2.9

Déniz et al. (2011) apresentam uma proposta de reconhecimento facial realizando uma amostragem uniforme para o cálculo do HOG na imagem da face. Além disso, o descritor HOG foi calculado considerando tamanhos diferentes para células, conseguindo codificar informações da face que contribuiriam na melhoria do reconhecimento. Como o cálculo do HOG é feito em várias escalas, isso torna o descritor muito grande e também foi realizada uma redução de dimensionalidade.

Li e Qi (2013) utilizaram o HOG, normalizado por uma norma ℓ_2 , como descritor da face e utilizaram uma codificação esparsa em grupo para realizar o reconhecimento. Nessa

¹Esses valores correspondem a uma representação de uma pessoa na imagem. Tais valores podem variar a depender do objeto, de acordo com Dalal e Triggs (2005)

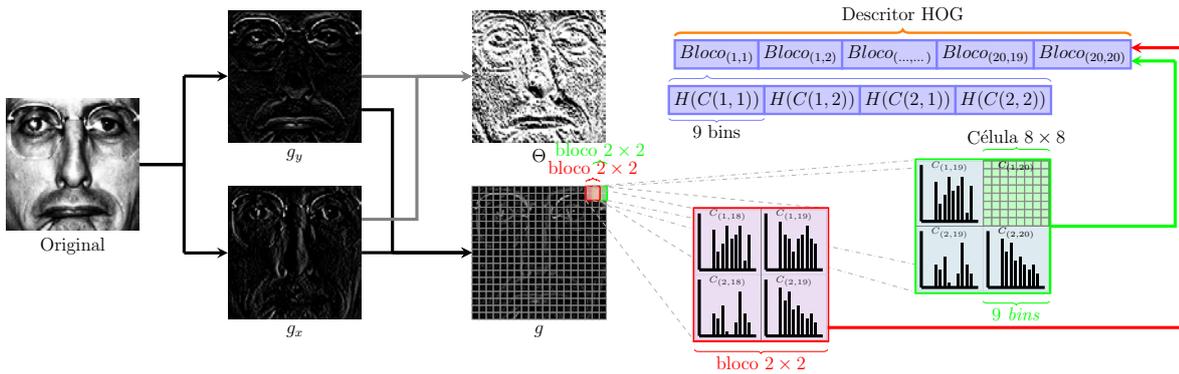


Figura 2.9: Extração do *Histogram of oriented gradients*. A imagem Original em escala de cinza é filtrada duas vezes, em x e em y gerando g_x e g_y ; com esses valores são calculados a magnitude g e a orientação θ ; a magnitude g é dividida em células de 8×8 e essas células são agrupadas em blocos sobrepostos de 2×2 (em vermelho e em verde); é feito o histograma da magnitude g agrupado por 9 bins da orientação θ ; a junção de todos os histogramas de todos os blocos compõem o HOG.

abordagem, as amostras de treinamento da mesma classe são reunidas para formar um subdicionário e um dicionário estruturado em grupo é obtido concentrando subdicionários de todas as classes.

Uma outra utilização do HOG foi o *locality sensitive histograms of oriented gradients* (LSHOG) proposto por Li e Huo (2016) como um novo descritor que é um HOG sensível à localidade e aplicado ao reconhecimento facial. O LSHOG proposto calcula um histograma de orientações de gradiente em cada pixel criando um vetor de tamanho 9 (*bins*). Em cada dimensão desse vetor será armazenado o somatório das direções dos gradientes calculados em todo os pixels da imagem em cada *bin* e multiplicado esse somatório por um parâmetro sensível à localidade. Esse parâmetro pode fazer com que o valor da direção do gradiente decline exponencialmente em relação à distância entre a localização do pixel do valor da direção do gradiente e a localização do pixel onde está sendo calculando o histograma.

2.2.4 Dual-cross patterns

O descritor *dual-cross patterns* (DCP) foi proposto por Ding et al. (2016) como um descritor focado na estrutura da textura da face tendo um custo computacional equivalente ao dobro do LBP. O DCP apresenta robustez à variação de pose e expressão. O descritor DCP realiza uma amostragem local considerando as direções onde são encontradas as informações mais relevantes da face que é a configuração dos componentes faciais e a forma de cada componente facial.

Após uma normalização geométrica da imagem do face, as partes centrais de vários componentes faciais, isto é, sobrancelhas, olhos, nariz e boca, se estendem horizontalmente ou verticalmente, enquanto as extremidades desses componentes faciais convergem

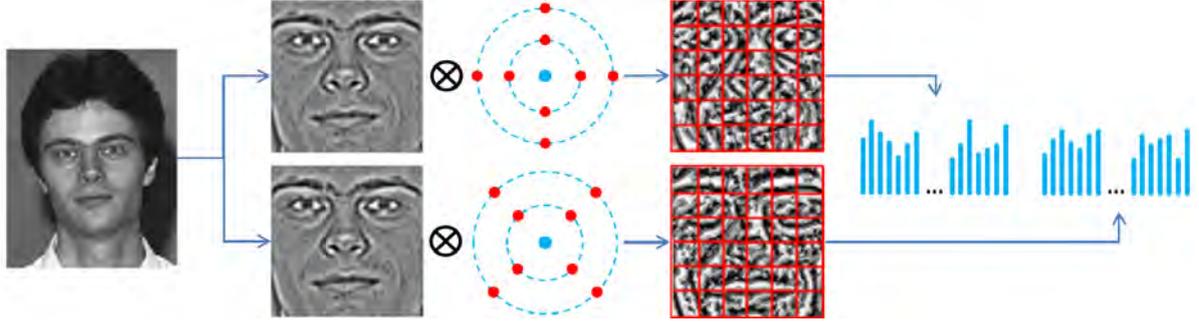


Figura 2.10: Representação do rosto usando DCP. A imagem de face normalizada é codificada pelos dois encoders cruzados, respectivamente. A concatenação dos histogramas regionais do código DCP forma a representação de face baseada em DCP. Figura extraída de (DING et al., 2016).

em direções diagonais (aproximadamente $\pi/4$ e $3\pi/4$). Levando isso em consideração o DCP definiu oito direções para realizar a amostragem: $0, \pi/4, \pi/2, 3\pi/4, \pi, 5\pi/4, 3\pi/2$ e $5\pi/4$. Sendo O o pixel central, dois valores de raios são definidos: o interno R_{in} e o externo R_{ex} . Em cada uma das direções são extraídas duas amostras, uma em cada raio, gerando dois conjuntos de 8 amostras: o conjunto $\{A_0, A_1, A_2, A_3, A_4, A_5, A_6$ e $A_7\}$ são as amostras extraídas em R_{in} e as amostras $\{B_0, B_1, B_2, B_3, B_4, B_5, B_6$ e $B_7\}$ extraídas em R_{ex} . Para realizar o cálculo do DCP, são considerados os valores, em escala de cinza, das amostras em R_{in} , chamadas de I_{A_i} , das amostras em R_{ex} , chamadas de I_{B_i} e do valor, em escala de cinza, do ponto central I_O conforme a seguir:

$$DCP_i = S(I_{A_i} - I_O) \times 2 + S(I_{B_i} - I_{A_i}), 0 \leq i \leq 7, \quad (2.5)$$

$$S(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{caso contrário} \end{cases} \quad (2.6)$$

Sendo 4 valores possíveis para cada direção, considerando as oito direções, o DCP pode gerar $4^8 = 65.536$ valores. Para diminuir esse valor, Ding et al. agruparam as oito direções em dois grupos, reduzindo para $4^4 \times 2 = 512$. Os dois grupos de direções foram chamados de DCP-1 e DCP-2 e o cálculo do descritor é obtido após o cálculo dos DCP-1 e DCP-2 em cada pixel da imagem, gerando dois mapas. Esses mapas são divididos em regiões não sobrepostas e é realizado o histograma dos DCP em cada região e depois esses histogramas são concatenados. O cálculo do descritor está ilustrado na Fig. 2.10.

2.2.5 SURF

O SURF é composto por um detector de pontos de interesse e um descritor dos pontos detectados, desenvolvido por Bay, Tuytelaars e Gool (2006). Os pontos de interesse do SURF possuem características que são invariantes à deslocamento, rotação e escalamento e parcialmente invariantes à iluminação e transformações afins.

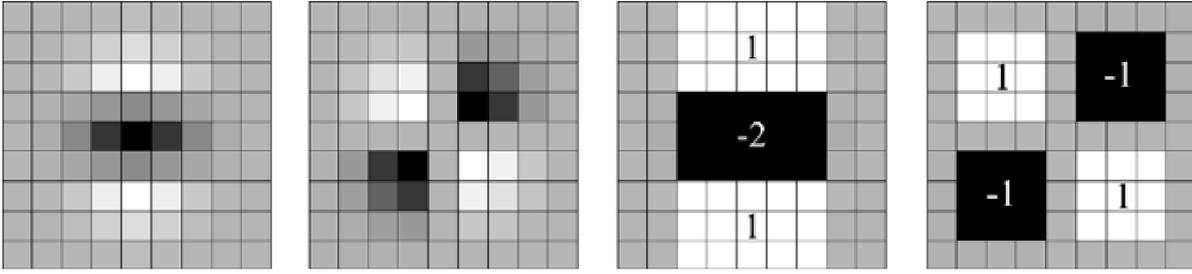


Figura 2.11: Da esquerda para a direita: Derivadas parciais gaussianas de segunda ordem (discretizadas e recortadas), com $\sigma = 1.2$, na direção y e direção xy , e aproximações usando filtros de caixa. As regiões cinza são iguais a zero. Figura extraída de (BAY; TUYTELAARS; GOOL, 2006).

Detector de pontos de interesse Para a detecção dos pontos de interesse é utilizada uma matriz Hessiana. Dado um ponto $P = (x, y)$ em uma imagem I , a matriz Hessiana no ponto $H(P, \sigma)$ em P e escala σ , é definida da seguinte forma:

$$H(P, \sigma) = \begin{bmatrix} L_{xx}(P, \sigma) & L_{xy}(P, \sigma) \\ L_{xy}(P, \sigma) & L_{yy}(P, \sigma) \end{bmatrix}, \quad (2.7)$$

onde $L_{xx}(P, \sigma)$ é a convolução da derivada da gaussiana de segunda ordem $\frac{\partial^2}{\partial x^2}g(\sigma)$ com a imagem I no ponto P e da mesma forma para $L_{xy}(P, \sigma)$ com $\frac{\partial^2}{\partial x \partial y}g(\sigma)$ e $L_{yy}(P, \sigma)$ com $\frac{\partial^2}{\partial y^2}g(\sigma)$.

A derivada da gaussiana de segunda ordem precisa ser discretizada antes de realizar a convolução com a imagem. No lugar das gaussianas são utilizadas aproximações, mostradas na Fig. 2.11, que oferecem resultados semelhantes na convolução, mas tornam o cálculo menos complexo computacionalmente, sendo rapidamente calculados usando imagens integrais.

A convolução da imagem com a derivada gaussiana de segunda ordem passa a ser chamada de D_{xx} , D_{yy} e D_{xy} e a determinante aproximada da matriz de Hessian é obtida da seguinte forma:

$$\det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2, \quad (2.8)$$

onde w é um peso usado para equilibrar a expressão para o determinante de Hessian sendo necessário para a conservação de energia entre os kernels gaussianos e os kernels gaussianos aproximados. Considerando o valor de $\sigma = 1.2$ tamanho do filtro 9×9 o valor de w será dado por:

$$w = \frac{|L_{xy}(1.2)|_F |D_{yy}(9)|_F}{|L_{yy}(1.2)|_F |D_{xy}(9)|_F} = 0.912 \dots \simeq 0.9, \quad (2.9)$$

onde $|x|_F$ é a norma de Frobenius. Dessa forma, o valor do peso está relacionado com a escala.

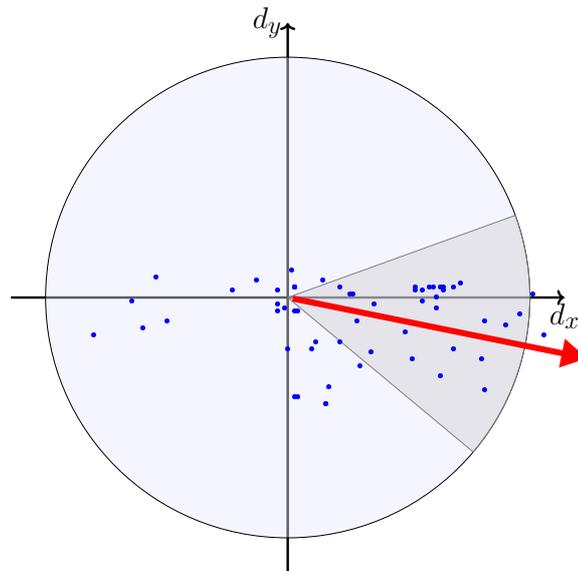


Figura 2.12: Atribuição de orientação: uma janela de orientação deslizante de tamanho $\frac{\pi}{3}$ detecta a orientação dominante das Haar wavelets ponderadas pela Gaussiana em cada ponto de amostra dentro de uma vizinhança circular ao redor do ponto de interesse. Adaptado de (BAY et al., 2008).

Os espaços de escala geralmente são implementados como pirâmides de imagens. As imagens são suavizadas repetidamente com um filtro gaussiano e posteriormente sub-amostrado para alcançar um nível mais alto da pirâmide. Como o SURF usa de filtros de caixa e imagens integrais, o espaço da escala é analisado por aumento do tamanho do filtro em vez de reduzir iterativamente o tamanho da imagem. A saída do filtro de 9×9 (mostrado na Fig. 2.11) é considerado como a camada de escala inicial, que é referida como escala $s = 1.2$ (correspondente a derivadas gaussianas com $\sigma = 1.2$).

As camadas seguintes são obtidas através da filtragem da imagem com máscaras gradualmente maiores, levando em consideração a natureza discreta das imagens integrais e a estrutura específica dos filtros utilizados, isso resulta em filtros de tamanho 9×9 , 15×15 , 21×21 , 27×27 , etc, fazendo com que a escala suba na mesma proporção ($\sigma = 3 \times 1.2 = 3.6 = s$)

A determinante aproximada da matriz hessiana é calculada a partir da geração de uma representação da imagem no ponto P em cada escala. Nessas representações é aplicada uma supressão não máxima em $3 \times 3 \times 3$ de vizinhança para encontrar o máximo. A localização e a escala (s) dos pontos do detector do SURF são obtidos pelos valores máximos (BROWN; LOWE, 2002).

Descritor O descritor do SURF é computado a partir da definição de uma orientação com base em informações de uma região circular em torno do ponto de interesse. Em seguida, é construída uma região quadrada alinhada à orientação definida pelo detector e o descritor SURF extraído essa região. Também existe uma versão com orientação vertical

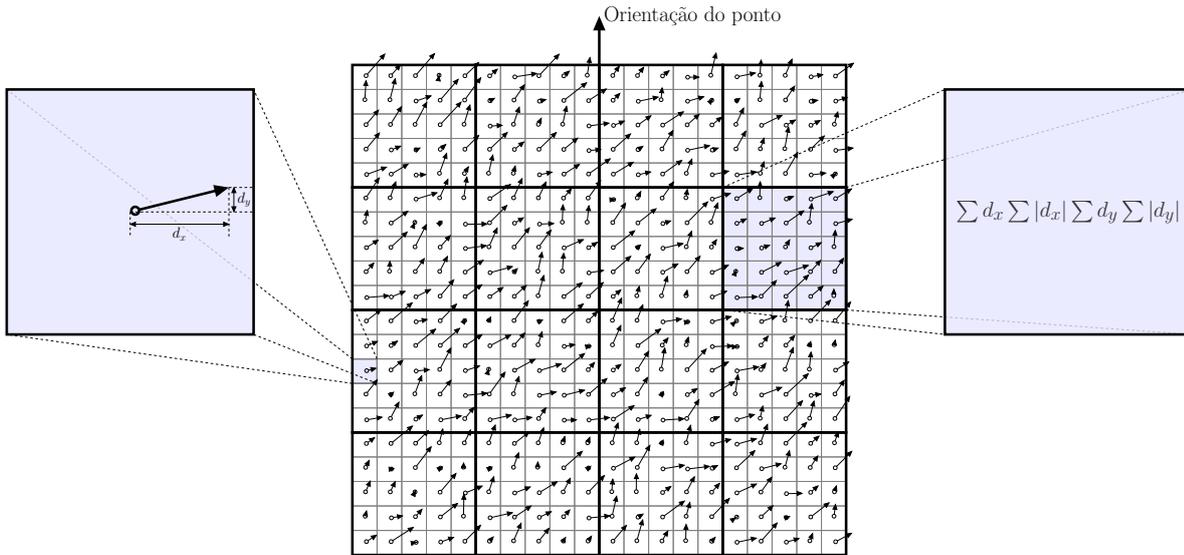


Figura 2.13: Para construir o descritor é colocado uma grade quadrada orientada dividida em sub-regiões 4×4 sobre o ponto de interesse (centro). Para cada sub-região, as respostas das Haar wavelets são calculadas a partir de amostras 5×5 . Para cada região, são coletadas as somas d_x , $|d_x|$, d_y , e $|d_y|$ (direita), calculadas em relação à orientação da grade (esquerda). Adaptado de (BAY et al., 2008).

previamente estabelecida do descritor (Upright SURF ou U-SURF) que não é invariante à rotação da imagem, melhor adaptado para aplicações onde a câmera permanece mais ou menos horizontal, porém mais rápido para calcular e mantendo uma robustez para uma rotação de cerca de $\pm 15^\circ$.

Para ser invariante à rotação, filtros de Haar wavelet são calculados nas direções x e y , em uma vizinhança circular de raio $6s$ ao redor do ponto de interesse, com s sendo a escala em que o ponto de interesse foi detectado. Uma vez que as Haar wavelet são calculadas e ponderadas com uma Gaussiana ($\sigma = 2,5s$) centrado no ponto de interesse, os resultados são representados como vetores em um espaço com peso horizontal ao longo da abscissa e peso vertical ao longo da ordenada, como pode ser visto na Fig. 2.12. A orientação dominante é estimada pelo cálculo da soma de todos os resultados dentro de uma janela de orientação deslizante cobrindo um ângulo de $\frac{\pi}{3}$. Os resultados horizontais e verticais dentro da janela são somados e esses dois resultados somados então produzem um novo vetor. O vetor mais longo dá sua orientação ao ponto de interesse. O U-SURF ignora esta etapa.

Para a extração do descritor, o primeiro passo é construir uma região quadrada centrada em torno do ponto de interesse e orientada ao longo da orientação selecionada na seção anterior com o tamanho de $20s$. A região é regularmente dividida em sub-regiões menores de 4×4 . Isso preserva informações espaciais importantes. Para cada sub-região, são calculadas as respostas de Haar wavelet em 5×5 pontos de amostra espaçados regularmente. Por razões de simplicidade, chamamos d_x a resposta da Haar wavelet na

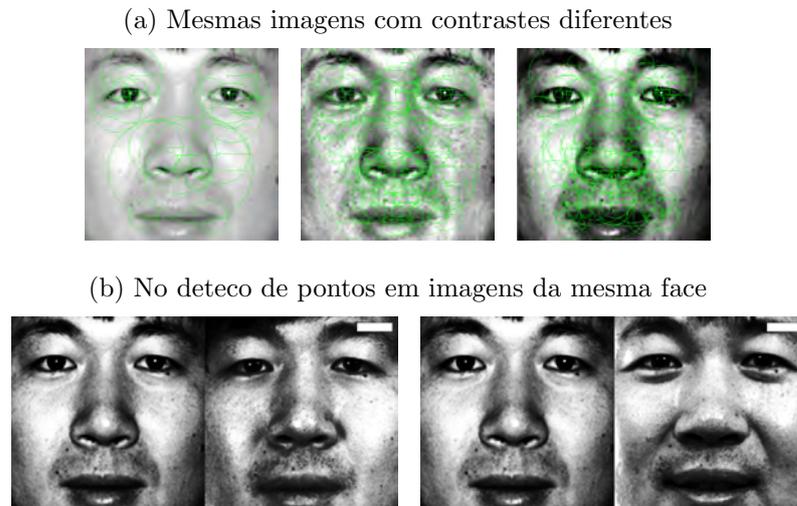


Figura 2.14: Variação da detecção de pontos de interesse do SURF devido à mudanças no contraste. (a) Mesma imagem com contraste diferentes e quantidade de pontos detectados diferentes. 24, 75 e 78 respectivamente. (b) Imagens da mesma face sem pontos em comum.

direção horizontal e d_y a resposta de Haar wavelet na direção vertical (tamanho de filtro 2s). “Horizontal” e “vertical” aqui são definidos em relação à orientação do ponto de interesse selecionado como pode ser visto na Fig. 2.13. Para aumentar a resistência contra deformações geométricas e erros de localização, as respostas d_x e d_y primeiro são ponderados com uma gaussiana ($\sigma = 3.3s$) centrado no ponto de interesse.

As respostas das Haar wavelets d_x e d_y são somadas em cada sub-região e formam um primeiro conjunto de entradas no vetor de características. Para trazer informações sobre a polaridade das mudanças de intensidade, extrai-se a soma dos valores absolutos das respostas, $|d_x|$ e $|d_y|$. Cada sub-região possui, portanto, um vetor descritor 4D v que é obtido por: $v = [\sum d_x, \sum |d_x|, \sum d_y, \sum |d_y|]$. A concatenação dos valores de v calculados em cada uma das sub-regiões 4×4 resulta em um descritor de dimensão 64.

Reconhecimento facial com SURF Alguns trabalhos utilizaram o SURF para reconhecimento facial, como por exemplo em Du, Su e Cai (2009) e Anand e Shah (2016) porém, o detector de pontos de interesse do SURF é muito sensível ao contraste. Essa



Figura 2.15: Detecção de pontos pelo SURF em locais afastados mas dados como o mesmo ponto.

sensibilidade faz com que a quantidade de pontos detectados possa variar muito na análise da mesma imagem com contrastes diferentes, como pode ser visto na Fig 2.14 item (a). A sensibilidade é tão grande que objetos semelhantes podem não ter nenhum ponto detectado como apresentado na Fig 2.14 item (b).

Outro problema encontrado para a utilização do SURF em reconhecimento facial é o fato de que as correspondências dos pontos de interesse detectados não consideram a forma do objeto na imagem fazendo com que pontos espacialmente afastados sejam considerados como o mesmo ponto. Essa situação está representada na 2.15.

2.2.6 Descritor de covariância

Tuzel, Porikli e Meer (2006) apresentaram um novo descritor de região, aplicado em detecção de objetos e classificação de texturas. Esse descritor calcula a covariância de características dentro de uma região de interesse da imagem e o seu desempenho foi superior à outros métodos mostrando que a matriz de covariância lida com rotações e mudanças de iluminação.

Na abordagem de Tuzel, Porikli e Meer (2006) 9 característica foram utilizadas para descrever a imagem, a posição x do pixel, a posição y , os valores de R, G, B (na posição x, y) e os resultados da aplicação dos filtros $[-1 \ 0 \ 1]^T$ e $[-1 \ 2 \ -1]^T$. Esse conjunto de características pode ser descrito da seguinte forma:

$$F(x, y) = [x \ y \ R(x, y) \ G(x, y) \ B(x, y) \ \left| \frac{\partial I(x, y)}{\partial x} \right| \ \left| \frac{\partial I(x, y)}{\partial y} \right| \ \left| \frac{\partial^2 I(x, y)}{\partial x^2} \right| \ \left| \frac{\partial^2 I(x, y)}{\partial y^2} \right|], \quad (2.10)$$

onde I é a intensidade do pixel.

As matrizes de covariância são de baixa dimensão em comparação a outros descritores de região, como por exemplo HOG e o Haar wavelet. Por ser uma matriz simétrica a matriz de covariância terá apenas $(d^2 + d)/2$ valores diferente, onde d é a dimensão do vetor (quantidade de características). No caso das características utilizadas por Tuzel, Porikli e Meer (2006) a matriz de covariância terá uma dimensão de 9×9 mas apenas 72 valores diferente.

Enquanto que a variância é uma medida da distribuição dos valores de uma variável ao redor da sua média a covariância dá uma ideia da direção da relação linear entre duas variáveis. Considerando as duas variáveis Y_i e Z_i ($i = \{1, 2, 3, \dots, N\}$) podemos dizer que:

$$\text{cov}(Y, Z) = \sum_{i=1}^N (Y_i - \mu_y) (Z_i - \mu_z), \quad (2.11)$$

onde μ_y é a média da população de Y_i e μ_z é a média da população de Z_i .

Valores positivos de covariância mostra que as duas variáveis se movem para a mesma direção, ou direções opostas, quando o sinal é negativo. Quando o valor da covariância entre duas variáveis é zero quer dizer que essas variáveis são independentes. Apesar da covariância crescer com o poder do relacionamento entre as variáveis, um valor alto não

quer dizer que o relacionamento entre as variáveis seja grande pois depende da escala dos valores fazendo com que a correlação mostre a direção da relação linear, mas não a força.

2.3 CLASSIFICADORES

No processo de reconhecimento facial, os dados extraídos das imagens são enviados para um classificador a fim de se determinar a classe em que aquelas informações se enquadram melhor. O reconhecimento facial ocorre quando o classificador consegue distinguir as imagens da face de uma mesma pessoa das imagens de outras. O limite conceitual entre a extração de características e a classificação adequada é um pouco arbitrário, já que um extrator de característica ideal forneceria uma representação que tornaria trivial o trabalho do classificador e, por outro lado, um classificador onipotente não precisaria da ajuda de um extrator sofisticado para realizar a classificação (DUDA; HART; STORK, 2001).

Duas estratégias de classificação são utilizadas quando uma face-de-busca é passada para o sistema de classificação facial, uma delas é a classificação não-supervisionada e a outra é a classificação supervisionada.

2.3.1 Classificação não-supervisionada

Em uma classificação não-supervisionada as características extraídas das imagens da galeria são inseridas em um espaço vetorial, onde cada nova inserção registra um ponto nesse espaço fazendo com que o classificador construa grupos desses pontos. Sendo assim, as características de uma imagem face-de-busca são comparadas com os pontos inseridos nesse espaço vetorial para encontrar uma correspondência. A Figura 2.16 ilustra a classificação não-supervisionada do reconhecimento facial no item (a), onde S_i é uma distância vetorial entre o ponto que representa o vetor das características da imagem que representa a face-de-busca e o ponto que representa o vetor de característica da imagem face-alvo, r é correspondente a menor distância entre o ponto da face-de-busca e os pontos das face-alvo, e i do S_i mínimo indicando qual a face-alvo mais próxima da face-de-busca.

O item (b) da Figura 2.16 ilustra a classificação não-supervisionada da verificação facial, onde V_p é o vetor de características da imagem face-de-busca, V_t é o vetor de características da imagem face-alvo, T é o limiar que determina a distância mínima para considerar que os dois vetores são representações da mesma face e r é a resposta do classificador que será “sim” caso o valor da distância entre V_p e V_t , dada por $d(V_p, V_t)$, seja menor que o limiar ou “não” no caso contrário.

Podemos citar como uma estratégia de classificação não-supervisionada o *nearest neighbor* (NN), proposto no trabalho de Cover e Hart (1967), onde apresentam uma estratégia de classificação baseada na regra do vizinho mais próximo. Com essa regra uma face-de-busca será classificada com a mesma classe que foi associada as imagens da galeria que ficarem mais próximas no espaço vetorial. Esse classificador já foi utilizado em alguns trabalhos de reconhecimento de faces como em (HE et al., 2005) e em (WEINBERGER; BLITZER; SAUL, 2006).

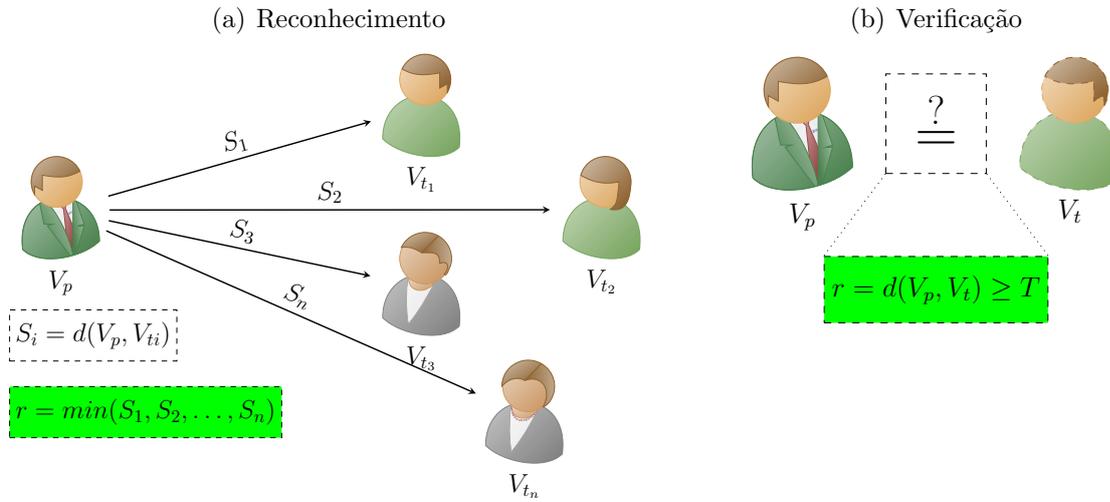


Figura 2.16: Classificação não-supervisionada. (a) Reconhecimento em uma classificação não-supervisionada onde S_i é uma distância vetorial, r é correspondente a mínima distância, e i indicando qual é a face mais próxima. (b) Verificação em uma classificação não-supervisionada onde V_p é o vetor de características da imagem face-de-busca, V_t é o vetor de características da imagem face-alvo, T é o limiar e r é a resposta do classificador que será “sim” caso o valor da distância $d(V_p, V_t)$ seja menor que o limiar ou “não” no caso contrário.

2.3.2 Classificação supervisionada

Em uma classificação supervisionada as características extraídas das imagens da galeria são apresentadas, juntamente com a identidade da face na imagem, para o classificador. O classificador irá construir um modelo a partir dos dados apresentados fazendo com que cada identidade seja uma classe e, com esse modelo, realizar previsões de classes para características extraídas de imagens que nunca foram apresentadas ao classificador.

A Figura 2.17 ilustra a classificação não-supervisionada do reconhecimento facial no item (a), onde V_j é uma concatenação do vetor características da imagem face-de-busca, V_p , com o vetor características da imagem face-alvo, V_{t_j} . Ψ é uma função não linear, W são os pesos, b é o Bias, T é o limiar e C será a resposta da função não linear para V_j . V_p será reconhecido caso $\sum_j C$ forma maior ou igual à 1 e a classe correspondente a V_p será o j que a $t_j(V_j)$ seja igual 1. No caso de $\sum_j C = 0$ quer dizer que V_p não é conhecido pelo classificador.

O item item (b) da Figura 2.16 ilustra a classificação supervisionada da verificação facial, onde V é uma concatenação do vetor características da imagem face-de-busca, V_p , com o vetor características da imagem face-alvo, V_t . ψ é uma função não linear, W são os pesos, b é o Bias, T é o limiar e $f(V)$ é a resposta do classificador que será “sim” caso o valor da função seja maior ou igual ao limiar ou “não” no caso contrário.

Uma estratégia de classificação supervisionada é a *support vector machine* (SVM) que foi introduzido por Cortes e Vapnik (1995) e apresenta por Cristianini e Shawe-Taylor

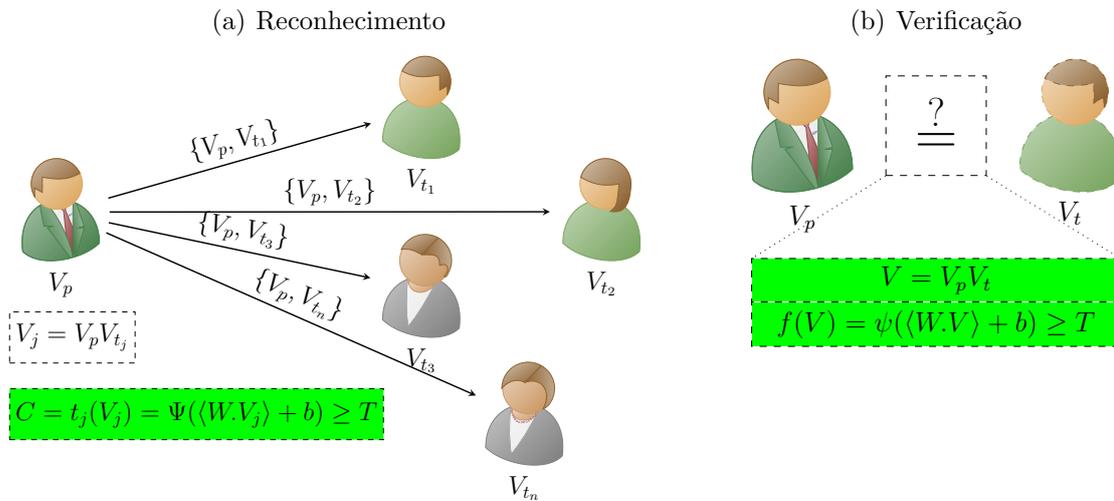


Figura 2.17: Classificação supervisionada. (a) Reconhecimento em uma classificação supervisionada onde V_j é uma concatenação dos vetore V_p , face-de-busca, e V_{t_j} , face-alvo, Ψ é uma função não linear, W são os pesos, b é o Baias, T é o limiar e C será a resposta da avaliação de V_j . V_p será reconhecido caso $\sum_j C \geq 1$ e a resposta do classificador será o j que a $t_j(V_j)$ seja igual 1. No caso de $\sum_j C = 0$ V_p não é conhecida. (b) Verificação em uma classificação supervisionada onde V é uma concatenação dos vetore V_p (características da imagem face-de-busca) e V_t (características da imagem face-alvo), ψ é uma função não linear, W são os pesos, b é o Baias, T é o limiar e $f(V)$ é a resposta do classificador que será “sim” caso o valor da função seja ≥ 0 ou “não” no caso contrário.

(2000) como sistemas de aprendizagem que usam um espaço de hipóteses de funções lineares construindo um hiperplano N-dimensional, treinado com um algoritmo de aprendizagem, derivado da teoria da aprendizagem estatística, para o reconhecimento de padrões ou estimativa de regressão.

No reconhecimento facial a SVM é treinada com uma galeria aprendendo que as imagens face-alvo de uma mesma pessoa devem permanecer em uma mesma classe. Após esse treinamento, quando uma face-de-busca for apresentada à SVM, essa irá apresentar como resposta a classe que a face-alvo seja mais parecida com a face-de-busca. Esse classificador já foi bastante utilizado em reconhecimento facial com pode ser visto em (HEISELE; HO; POGGIO, 2001).

O algoritmo de classificação *Linearly Approximated Sparse Representation-based Classification* (LASRC) foi proposto por Ortiz e Becker (2014) e usa regressão linear para realizar a seleção da amostra para minimização ℓ_1 , aproveitando assim a velocidade dos mínimos quadrados e a robustez das soluções dispersas, como o SRC.

Ortiz e Becker (2014) disponibilizaram uma implementação de um *toolbox* do Matlab chamada de *Matlab Face Recognition Toolbox*² e nessa implementação foram postos juntos vários classificadores com a intenção de realizar uma comparação entre eles.

²<http://enriqueortiz.com/wordpress/enriqueortiz/research/face-recognition/webscale-face-recognition/>

Dentre os classificadores comparados com o trabalho de Ortiz e Becker tem o trabalho de Wright et al. (2009) que apresentaram um algoritmo de classificação geral aplicando para o reconhecimento de objetos com base em uma representação dispersa calculada por minimização ℓ_1 o *Sparse Representation-based Classification* (SRC).

SRC segue a teoria da *Compressed Sensing* que diz a escolha precisa do espaço das características não é mais crítica (WRIGHT et al., 2009). Segundo essa teoria mesmo característica aleatórias contêm informações suficientes para recuperar a representação dispersa e, portanto, classificar corretamente qualquer face-de-busca. Para isso, a dimensão do espaço de características deve ser suficientemente grande para que a representação dispersa seja calculada corretamente.

O SRC representa a face-de-busca por todas as face-alvo, que é demorado quando o número de face-alvo é muito grande. Além disso, o SRC pode representar a face-de-busca de forma muito esparsa pelas face-alvo que são realmente partes da face-de-busca. Se essas face-alvo forem de uma classe diferente da classe da face-de-busca o SRC pode levar a uma classificação incorreta (ZHANG; YANG, 2010).

Aproveitando os benefícios das duas abordagens, uma combinação entre o NN e o SRC foi apresentada por Zhang e Yang (2010), o *K nearest neighbor based local sparse representation classifier* (KNN-SRC) que escolhe os vizinhos mais próximos de uma face-de-busca entre todas as face-alvo para representar a face-de-busca.

Outros trabalhos baseados na representação dispersa foram o de Figueiredo, Nowak e Wright (2007), que apresentam a *Gradient Projection for Sparse Reconstruction* (GPSR) e Gan e Xiao (2012) que mostram uma combinação de SRC com Homotopia (OSBORNE; PRESNELL; TURLACH, 2000).

Ortiz e Becker (2014) também comparou seu classificador com o *Linear Regression Classification* (LRC) que um conceito de que os padrões de uma mesma classe de objeto residem em um subespaço linear que foi aplicado em reconhecimento facial em (NASEEM; TOGNERI; BENNAMOUN, 2010). Foi desenvolvido um modelo linear representando uma face-de-busca como uma combinação linear de galerias específicas de classe e o problema inverso é resolvido usando o método de mínimos quadrados. A decisão do classificador é em favor da classe com o mínimo erro de reconstrução.

Outro classificador comparado com o de Ortiz e Becker (2014) foi o classificador *Locally-constrained Linear Coding* (LLC) que utiliza as restrições de localidade para projetar cada descritor em seu sistema de coordenadas locais e essas coordenadas projetadas são combinadas para gerar a representação final (WANG et al., 2010).

Mais um classificador comparado com o LASRC foi o trabalho de Shi et al. (2011) que utilizada uma estratégia de minimização ortonormal de ℓ_2 (MO-L2) para o problema de reconhecimento facial e mostram que essa é significativamente precisa, robusta e muito rápida. Esse MO-L2 também foi avaliado e comparado com o trabalho de Ortiz e Becker (2014).

2.4 DATA SETS

Para reconhecimento de faces, os data sets são compostos com pelo menos duas imagens de cada face para que o sistema tenha uma como face-de-busca e a outra como face-alvo.



Figura 2.18: Exemplos de imagens normalizadas de uma mesma face do data set FERET.

Alguns data sets fazem uma separação entre as imagens que devem ser utilizadas como conjunto de treinamento, testes e validação, já outros deixam para que essa divisão seja feita pelo próprio sistema de reconhecimento de face. O protocolo padrão de avaliação de sistemas de reconhecimento de face requer imagens de treinamento e as imagens de teste separados. O conjunto de treinamento é utilizado para que o classificador aprenda as classes enquanto que os conjuntos de teste são usados para testar o sistema de reconhecimento.

Em princípio, não deve haver sobreposição entre o conjunto de imagens de treinamento e o conjunto de imagens de teste. Idealmente, para garantir que o sistema não esteja ajustado a nenhuma condição específica, as imagens de treinamento e as imagens de teste devem ser originárias de fontes diferentes e independentes. Infelizmente nem sempre é possível seguir esse protocolo, fazendo com que, na maioria das vezes, um único conjunto de imagens com identidades conhecidas seja dividido em treino e teste (DATTA; DATTA; BANERJEE, 2016). A seguir serão apresentados os data sets públicos utilizados nesta dissertação.

2.4.1 Facial recognition technology

O *Facial Recognition Technology* (FERET) foi criado por Phillips et al. (1997) e esse data set estabelece uma grande base de dados de imagens faciais, as quais foram coletada de forma independente dos desenvolvedores do algoritmo além de definir um protocolo



Figura 2.19: Exemplos de imagens normalizadas de uma mesma face do data set IMM.



Figura 2.20: Exemplos de imagens normalizadas de uma mesma face do data set ORL.

discriminando as imagens que deveriam ser utilizadas para treino e as imagens para teste.

O data set FERET foi coletado em 15 sessões entre agosto de 1993 e julho de 1996 e contém 1564 conjuntos de imagens totalizando 14.126. Como o equipamento teve que ser remontado para cada sessão, houve uma pequena variação nas imagens coletadas em datas diferentes. Para alguns indivíduos, decorreram mais de dois anos entre as suas primeiras e últimas sessões. Esse lapso de tempo foi importante, pois permitiu que os pesquisadores estudassem, pela primeira vez, mudanças na aparência de uma pessoa ao longo desse tempo. Alguns exemplos de imagens do data set FERET podem ser vistos na Fig. 2.18.

2.4.2 The IMM face database

Nordstrøm et al. (2004) produziram no instituto de Informática e Modelagem Matemática da Universidade Técnica da Dinamarca um data set com 240 imagens de 40 indivíduos. Esse data set foi construído com imagens de 7 mulheres e 33 homens sem a utilização de óculos e teve algumas de suas imagens analisadas em (STEGMANN; ERSBØLL; LARSEN, 2003). O IMM data set tem 6 imagens para cada pessoa seguindo as regras da tabela 2.1 e alguns exemplos dessas imagens podem ser vistos na Fig. 2.19.

Tabela 2.1: Regras das imagem do data set IMM para cada pessoa.

Imagem	Orientação da Face			Expressão			Luz	
	Fron- tal	30° direita	30° es- querda	Neu- tra	Feliz	Ar- bitrária	Difusa	Lado
1	✓			✓			✓	
2	✓				✓		✓	
3		✓		✓			✓	
4			✓	✓			✓	
5	✓			✓				✓
6	✓					✓	✓	



Figura 2.21: Exemplos de imagens normalizadas de uma mesma face do data set Yale na primeira linha e do Extended Yale Face Database B na segunda linha.

2.4.3 The ORL database of faces

O ORL data set foi apresentado primeiro no trabalho de Samaria e Harter (1994), e consiste em 400 imagens divididas em 10 imagem por pessoa. Os participantes são funcionários da Olivetti Research Ltd. ou estudantes da Universidade de Cambridge e têm idades variando entre 18 e 81 anos, sendo a maioria dos indivíduos com idade entre 20 e 35. São 4 mulheres e 36 homens e foram convidados a encarar a câmera sem restrições à expressão, apenas tendo seus movimentos laterais e inclinação limitados. A maioria dos participantes teve sua imagem capturada em diferentes momentos e com diferentes condições de iluminação, mas sempre contra um fundo escuro sendo que algumas pessoas estavam com óculos durante as fotos. Alguns exemplos de imagens desse data set podem ser vistos na Fig. 2.20.

2.4.4 Yale data sets

Na análise experimental do extrator proposto, foram utilizados dois data sets da Yale: O primeiro, o Yale Face Database ³ contém 165 imagens em escala de cinza em formato GIF de 15 indivíduos. Existem 11 imagens por indivíduo sendo uma para cada expressão facial, ou configuração, a seguir: luz central, com óculos, feliz, luz pela esquerda, sem óculos, normal, luz pela direita, triste, com sono, surpresa e piscando um dos olhos. Alguns exemplos de imagens desse data set podem ser vistos na primeira linha da Fig. 2.21.

O segundo foi o Extended Yale Face Database B que contém 38 indivíduos com 9 poses e 64 condições de iluminação. Nesse segundo data set, a variação de iluminação é bem mais acentuada e as imagens são divididas em cinco subconjuntos de acordo com o ângulo entre a direção da fonte de luz e o eixo central da câmera (12°, 25°, 50°, 77°, 90°). Alguns exemplos de imagens desse data set podem ser vistos na segunda linha da Fig. 2.21.

³<http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html>



Figura 2.22: Exemplos de imagens normalizadas de uma mesma face do data set PF83LFW.

2.4.5 PubFig83+LFW (PF83LFW)

O PF83LFW ⁴ data set é uma junção entre os data sets PubFig83⁵ e LFW⁶ para formar um novo conjunto de dados de referência para a identificação de face de “universo aberto” (BECKER; ORTIZ, 2013). Esse data set é baseado em cenários realistas onde ocorrem buscas automáticas de pessoas em fotos da web, ou amigos e familiares, que possam estar marcadas em álbuns de fotos pessoais. O propósito desse data set é permitir que algoritmos encontrem e identifiquem alguns indivíduos enquanto ignora todos as distrações como as paisagem ou faces de outras pessoas. Isso simula muitas das aplicações do mundo real onde o reconhecimento facial precisa ignorar muitos rostos de fundo que aparecem nas fotos que não são relevantes para o usuário.

PF83LFW tem 13.002 rostos representando 83 indivíduos de PubFig83, divididos em 2/3 de treinamento (8.720 faces) e 1/3 de teste (4.282 faces). Do LFW, 12.066 rostos que representam mais de 5.000 imagens são usados como um conjunto de distração. Alguns exemplos de imagens desse data set podem ser vistos na Fig. 2.22.

2.5 RELAÇÃO COM O TRABALHO PROPOSTO

Com os resultados superiores do trabalho de Tuzel, Porikli e Meer (2006) a partir da utilização da matriz de covariância para descrever múltiplas características da imagem, decidiu-se investigar a sua forma normalizada – a matriz de correlação. A hipótese é que a normalização interna dos descritores, ao invés da normalização vetorial (*e.g.*, normal ℓ_1 , norma ℓ_2 , norma ℓ_2 -histerese, etc), poderia trazer maior discriminabilidade para a representação proposta. Dessa forma, iniciaram-se algumas análises experimentais para a escolha do melhor conjunto de características a serem descritas pela matriz de correlação (ver Capítulo 4). As matrizes de correlação, inicialmente, eram calculadas considerando as características extraídas da face-de-busca de forma holística (toda a imagem da face) e seus resultados foram comparados com as matrizes de correlação das faces-alvo, por meio de métricas de distância (método não-supervisionado). Depois

⁴<http://www.brianbecker.com/blog/research/pubfig83-lfw-dataset/>

⁵<http://vision.seas.harvard.edu/pubfig83/>

⁶<http://vis-www.cs.umass.edu/lfw/>

disso, outro experimento foi analisado, utilizando a imagem da face-de-busca e faces-alvo divididas em regiões imagéticas (*patches*); as matrizes de correlação passaram a ser calculadas em cada região (*patch*). Nessa última configuração experimental, cada imagem passou a ser representada por um conjunto de matrizes de correlação de acordo com a quantidade de regiões divididas da imagem. Com isso, para realizar a comparação entre as imagens, os valores resultantes das distâncias entre as matrizes de correlação de regiões correspondentes das duas imagens comparadas eram somados e os valores-somatórios entre a face-de-busca e as faces-alvo ordenados. O valor reconhecido da face-alvo na galeria deveria ser o menor valor após a ordenação (*TOP 1*). Para obter o conjunto de características mais discriminativo foi necessário utilizar um método de busca e otimização no espaço de características baseado em algoritmo genético. O objetivo da otimização foi avaliar se poderia haver sub-conjuntos de características que pudesse representar melhor o data set utilizado para validação. Esse processo demonstrou que mesmo o melhor sub-conjunto de características não conseguia ultrapassar o estado-da-arte (DING et al., 2016).

Com base no trabalho de Bay, Tuytelaars e Gool (2006), que apresentam um descritor de tamanho 64, calculado de forma esparsa em pontos de interesse na imagem, o COHAWES foi proposto. A hipótese era que a utilização densa do descritor poderia trazer melhores resultados. Decidiu-se utilizar uma metodologia de avaliação com base em (ORTIZ; BECKER, 2014) com o objetivo de verificar o poder discriminativo do COHAWES a partir da utilização de diversos classificadores supervisionados e não-supervisionados (NN, LLC, KNN-SRC, ℓ_2 , LRC, SVM, LASRC, SRC-GPSR, SRC-Homotopy) e comparar os resultados com outros extratores/descriptores de características: *Gabor wavelets* (WISKOTT et al., 1997), LBP (OJALA; PIETIKÄINEN; HARWOOD, 1996), HOG (DALAL; TRIGGS, 2005) e DCP (DING et al., 2016). Este último extrator foi acrescido à metodologia utilizada em (ORTIZ; BECKER, 2014). Um objetivo adicional era comparar também o desempenho da correlação com a covariância, dentro dessa última metodologia adotada. O desempenho comparativo entre os diversos extratores e classificadores foi obtido sobre 8 data sets: FERET (sub-conjuntos FB, DUP1 e DUP2) (PHILLIPS et al., 1997), IMM (NORDSTRØM et al., 2004), ORL (SAMARIA; HARTER, 1994), Yale (BELHUMEUR; HESPANHA; KRIEGMAN, 1997), extended Yale face database B (GEORGHIADES; BELHUMEUR; KRIEGMAN, 2001), PF83LFW (BECKER; ORTIZ, 2013). O COHAWES obteve um resultado médio superior considerando todos os cenários testas, bem como obteve resultado superior à variação do COHAWES utilizando matriz de covariância em 7 dos 8 data sets.

CORRELATION OF HAAR WAVELET SUMS

3.1	Visão geral do COHAWES	33
3.2	Descrevendo cada pixel da imagem	35
3.3	Divisão da imagem em regiões	38
3.4	Obtendo o vetor de características	38
3.4.1	Correlação dentro de cada região	39
3.4.2	Concatenação do vetor de características	40
3.5	Considerações finais	41

Para a tarefa de reconhecimento facial, o COHAWES é calculado numa imagem que já tenha passado pelo processo de normalização, o qual é realizada da seguinte forma: (i) a localização dos olhos e da boca são obtidas a partir de um detector de pontos fiduciais; (ii) com os olhos detectados, a imagem da face é rotacionada para que os olhos fiquem alinhados ao eixo horizontal; (iii) com a localização da boca, a imagem rotacionada da face é recortada e redimensionada para um tamanho de 160×160 pixels. Todo este processo é feito a partir da imagem de entrada convertida para escala de cinza. Este processo de normalização é importante para auxiliar o sistema de reconhecimento facial a obter uma certa invariância nas variações de iluminação e rotação. É esperado que o COHAWES também apresente um certo nível de invariância, mas isso não foi avaliado neste trabalho.

3.1 VISÃO GERAL DO COHAWES

O cálculo do COHAWES é realizado em 4 etapas, ilustradas na Fig. 3.1. Na primeira etapa (i), é calculado um vetor de tamanho 64 para cada pixel da imagem. Esse vetor tem seus valores obtidos pelos somatórios dos gradientes horizontais (d_x) e verticais (d_y) da aplicação dos filtros Haar wavelet. Para se obter esses somatórios uma região 20×20 é delimitada ao redor de cada pixel e essa região é dividida de forma regular em 16

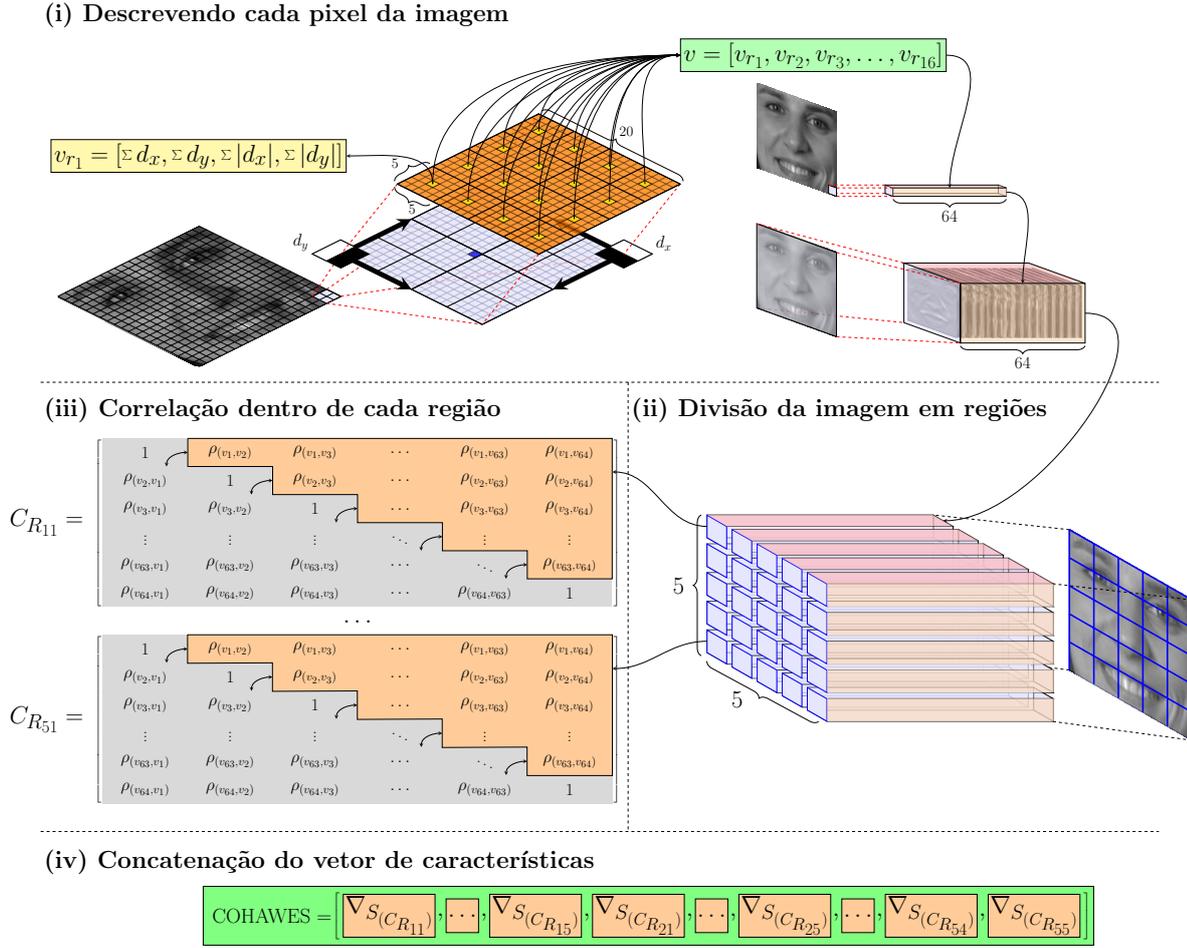


Figura 3.1: Visão geral do COHAWES. (i) Aplicação dos filtros Haar wavelet, na vertical (d_y) e na horizontal (d_x), definida em uma região 20×20 para cada pixel e o cálculo de $\sum d_x$, $\sum d_y$, $\sum |d_x|$, $\sum |d_y|$ nas 16 sub-regiões. Vetorizando os valores obtidos em (i), obtemos o vetor v , de tamanho 64, em cada pixel; (ii) 25 grupos de vetores v são extraídos de acordo com a segmentação da imagem original; (iii) cálculo da matriz de correlação para cada grupo de vetores; (iv) vetorização do COHAWES a partir da concatenação dos valores acima da diagonal principal das matrizes de correlação (∇_S).

sub-regiões de tamanho 5×5 onde, em cada sub-região são calculados os somatórios $\sum d_x$, $\sum d_y$, $\sum |d_x|$, $\sum |d_y|$. Como em cada sub-região são obtidos 4 valores, e são 16 sub-regiões, a concatenação dos valores obtidos nas sub-regiões será um vetor v de tamanho 64. Na etapa (ii), o conjunto de vetores de tamanho 64 são agrupados em 5×5 regiões, de acordo com a divisão da imagem. Na etapa (iii), cada uma das 64 características são consideradas variáveis e, a partir destas, são calculadas as matrizes de correção para cada grupo de vetores definidos pela divisão da imagem, onde $C_{R_{11}}$ representa a matriz de correlação da região da linha 1 coluna 1. Devido à característica simétrica da matriz de correlação, na etapa (iv), os valores da diagonal principal e da triangular inferior são

desconsiderados, e apenas os valores acima da diagonal principal (representados por ∇_S) das matrizes de correlação de cada região são considerados na vetorização do COHAWES. Os passos gerais para o cálculo do COHAWES é ilustrado em Alg. 3.1.

Algoritmo 3.1 Pseudocódigo do COHAWES

```

1:  /*
2:  Entrada:
3:      imagem = imagem da face normalizada
4:  Saída:
5:      COHAWES = descritor proposto
6:  Parâmetros:
7:      largura: largura da imagem em pixels (160)
8:      altura: altura da imagem em pixels (160)
9:      regioes: regiões as quais a matriz dos somatórios dos gradientes de Haar
              wavelet (bloco) será dividida (5x5)
10: */
11: //(i) Descrevendo cada pixel da imagem
12: imagemDx = filtroHaarWaveletDx(imagem)
13: imagemDy = filtroHaarWaveletDy(imagem)
14: bloco = zeros(largura, altura, 64)
15: para l de 1 ate largura faca
16:     para a de 1 ate altura faca
17:         v = [];
18:         //obtem vizinhança do pixel
19:         para vizLargura de 1 ate 4 faca
20:             para vizAltura de 1 ate 4 faca
21:                 somaDx, somaDy, somaAbsDx, somaAbsDy = 0
22:                 para vizRegiaoLargura de 1 ate 5 faca
23:                     para vizRegiaoAltura de 1 ate 5 faca
24:                         x = l-((3-vizLargura)*5)+vizRegiaoLargura
25:                         x = max(min(x, largura),0)
26:                         y = a-((3-vizAltura)*5)+vizRegiaoAltura
27:                         y = max(min(y, altura),0)
28:                         somaDx += imagemDx(x,y)
29:                         somaDy += imagemDy(x,y)
30:                         somaAbsDx += abs(imagemDx(x,y))
31:                         somaAbsDy += abs(imagemDy(x,y))
32:                     fimpara
33:                 fimpara
34:                 v = [v, somaDx, somaDy, somaAbsDx, somaAbsDy];
35:             fimpara
36:         fimpara
37:         bloco(l,a) = v;
38:     fimpara
39: fimpara
40: COHAWES = [];
41: //(ii) Divisão da imagem em regiões
42: para subbloco em dividirBlocoRegioes(bloco, regioes)
43:     //(iii) Correlação dentro de cada região
44:     Cr = correlacao(subbloco)
45:     TS = vetorizarDadosNaoRedundantes(Cr)
46:     //(iv) Concatenação do vetor de características
47:     COHAWES = [COHAWES, TS];
48: fimpara

```

3.2 DESCREVENDO CADA PIXEL DA IMAGEM

Em cada pixel da imagem, é definida uma região de tamanho 20×20 pixels e nessa região são aplicados 2 filtros Haar wavelet, um na horizontal e outro na vertical. Esses filtros podem ser vistos na Fig. 3.2. Cada filtro tem um tamanho 15×15 e para a sua aplicação

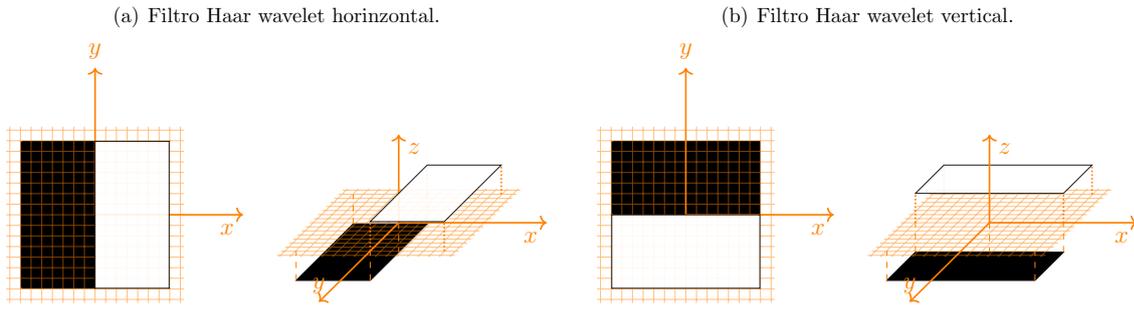


Figura 3.2: Filtros Haar wavelet de tamanho 15×15 . (a) Filtros Haar wavelet para calcular as respostas na direção x (b) Filtros Haar wavelet para calcular as respostas na direção y . As partes escuras têm o peso -1 e as partes claras $+1$.

ele é posicionado de forma que ocupe 7 linhas para cima e para baixo, 7 colunas para esquerda e para direita, do pixel central, fazendo com que alguns pixels fiquem na região clara do filtro e outros na região escura. Os pixels que ficam na região escura terão seus valores somados e esse total será subtraído do total dos pixels que ficam na região clara, gerando assim o gradiente de Haar. Quando esse gradiente é obtido pela aplicação do filtro vertical chamamos de d_y e quando é obtido pelo filtro horizontal chamamos de d_x .

Após o cálculo dos valores de d_x e d_y , a região de 20×20 é sub-dividida em 16 regiões de 5×5 . Em cada uma dessas sub-regiões, são calculados os somatórios de d_x e d_y , bem como os somatórios dos seus valores absolutos $|d_x|$ e $|d_y|$, gerando um vetor v de dimensão 4 que pode ser representado por $v = [\sum d_x, \sum |d_x|, \sum d_y, \sum |d_y|]$. Cada somatório informa o quanto de um determinado padrão de intensidade dos pixels prevalece na região. Valores altos em $\sum d_x$ e $\sum d_y$ indicam uma mudança no gradiente de intensidade dos pixels nas

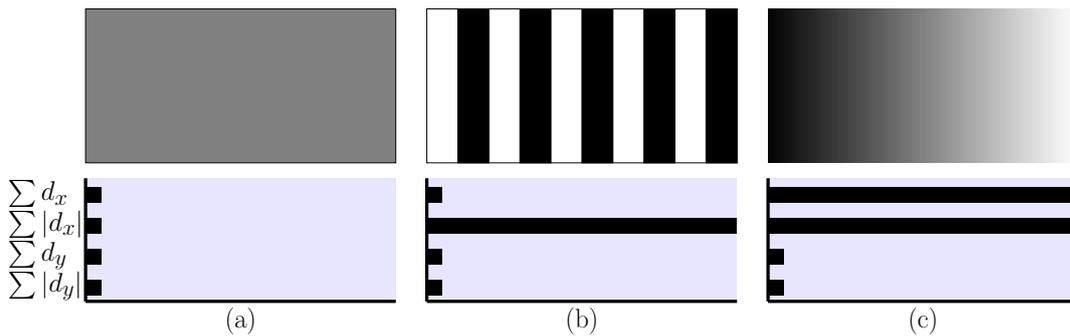


Figura 3.3: Os somatórios calculados em uma região representam o padrão de intensidade dos pixels. (a) Em uma região homogênea, todos os valores dos somatórios são relativamente baixos. (b) Em uma região que apresenta frequências na direção x , o valor de $\sum |d_x|$ é alto, mas todos os outros permanecem baixos. (c) A intensidade vai aumentando gradualmente na direção x , ambos os valores $\sum d_x$ e $\sum |d_x|$ são altos. Adaptado de (BAY et al., 2008).

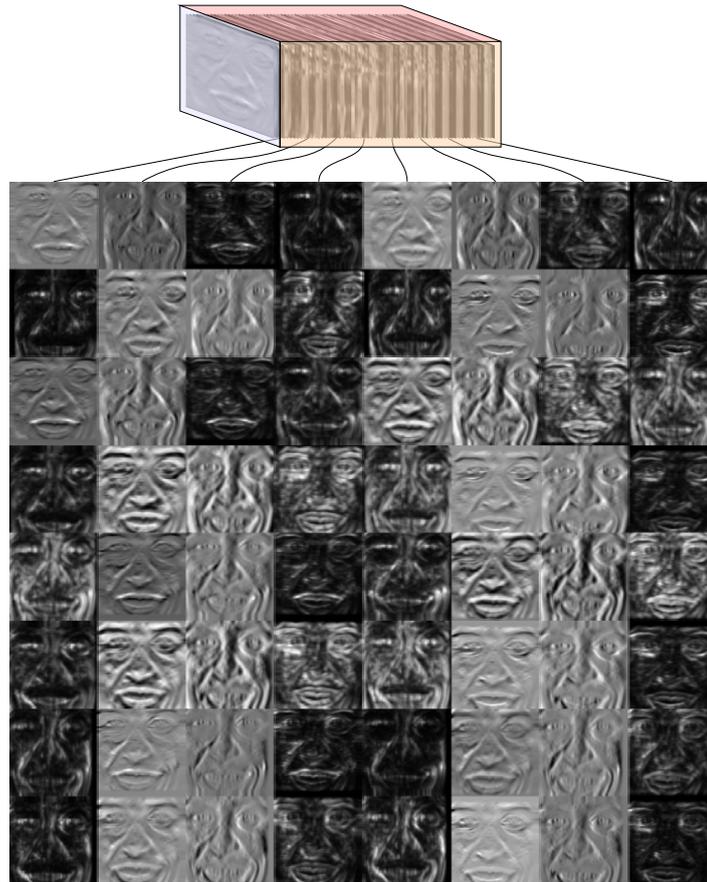


Figura 3.4: Características extraídas dos somatórios das Haar wavelet. São produzidas 64 novas matrizes do mesmo tamanho da imagem.

direções horizontal e vertical, respectivamente, enquanto que valores altos em $\sum |d_x|$ e $\sum |d_y|$ indicam presença de altas frequências de um determinado pixel na direção x e y , respectivamente.

A Figura 3.3 ilustra a influência dos valores de intensidade dos pixels nos somatórios onde em (a) são exibidos os valores dos somatórios todos baixos quando a imagem analisada tem valores de intensidade uniformes para os pixels; (b) apresenta um exemplo de imagem com valores de intensidade com frequências na horizontal e a influência disso nos somatórios fazendo com que o $\sum |d_x|$ tenha um valor maior do que os demais; (c) apresenta um exemplo de imagem com valores de intensidade com uma sequência de tons contínua na horizontal e a influência disso nos somatórios, fazendo com que o $\sum d_x$ e o $\sum |d_x|$ tenham um valor maior do que os demais.

Calculando os somatórios em cada uma das 16 regiões será gerado um vetor de tamanho 64 que irá descrever o pixel. Esse cálculo se repete para todos os pixels da imagem e está representado no item (i) da Fig. 3.1. Ao final desse processo, são produzidas 64 novas matrizes do mesmo tamanho da imagem que são ilustradas individualmente na Fig. 3.4.

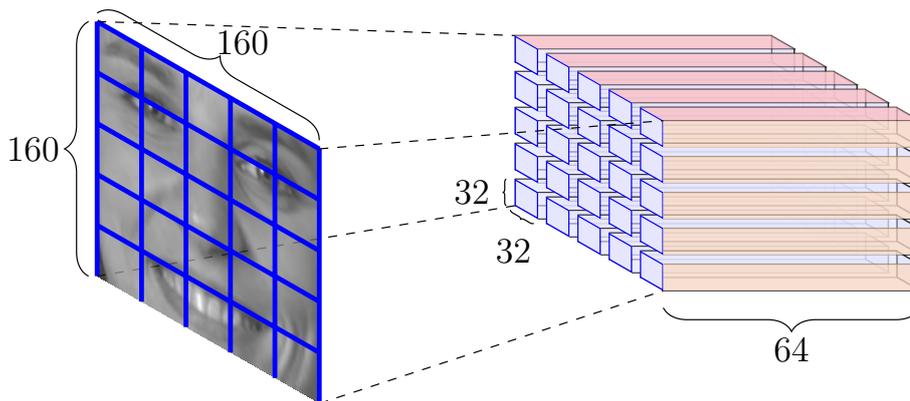


Figura 3.5: O bloco de vetores v é dividido em grupos seguindo uma divisão da imagem original.

3.3 DIVISÃO DA IMAGEM EM REGIÕES

Tomando a imagem original como referência, o conjunto de vetores v , que descreve cada pixel da imagem, é dividido em grupos. A escolha da quantidade e tamanho dos grupos é parametrizada permitindo que essa divisão possa ser facilmente modificada para o processo de extração do COHAWES. Esse processo de divisão está representado no item (ii) da Fig. 3.1.

Na avaliação da precisão do COHAWES, discutida no Capítulo 4, foi aplicada uma divisão uniforme em 25 regiões sendo isso feito em imagens normalizadas de 160×160 pixels. Os grupos de vetores são organizados em 5×5 e em cada grupo haverá $32 \times 32 \times 64$ valores. Essa divisão dos grupos pode ser vista na Fig 3.5. Além da divisão 5×5 pixels, outras divisões uniformes estão disponíveis pela parametrização e são representadas na Fig 3.6.

3.4 OBTENDO O VETOR DE CARACTERÍSTICAS

Após a divisão do bloco de vetores em grupos, cada grupo irá conter informações sobre os pixels, e informações sobre a vizinhança desses pixels, da região da imagem que originou o grupo. São 64 características e nessa etapa do COHAWES, essas características serão

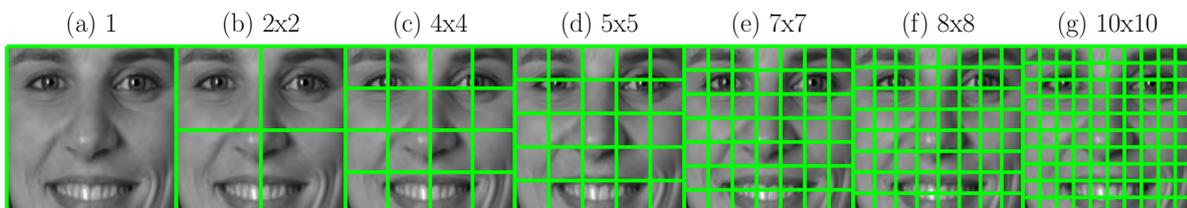


Figura 3.6: Divisões avaliadas durante o processo de construção do COHAWES.(a) Imagem inteira; (b) 2×2 ; (c) 4×4 ; (d) 5×5 ; (e) 7×7 ; (f) 8×8 ; (g) 10×10 .

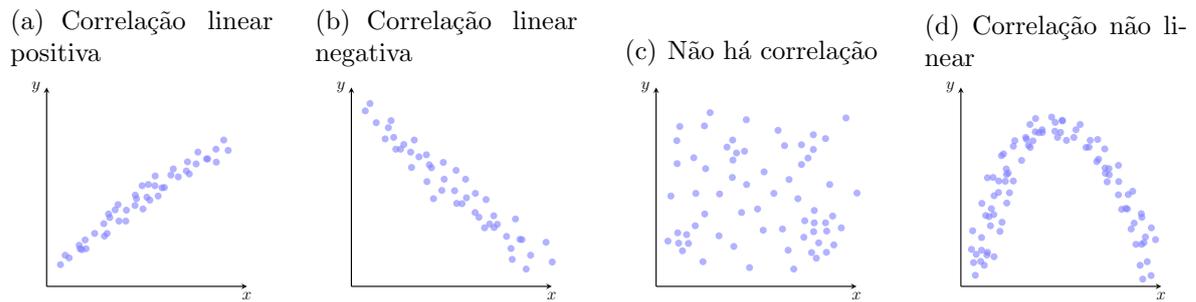


Figura 3.7: Diagramas de dispersão. (a) Os valores de y tende a aumentar quando o valor de x cresce. (b) Os valores de y tende a decrescer quando o valor de x cresce. (c) Não apresenta correlação entre o valor de y o valor de x . (d) Existe uma correlação entre o valor de y o valor de x mas essa correlação não é linear. Adaptado de (LARSON; FARBER, 2010)

analisadas dentro dos grupos; isso acontece a partir do cálculo da matriz de correlação entre os vetores de características de cada pixel, gerando assim uma matriz simétrica para cada grupo.

3.4.1 Correlação dentro de cada região

Um diagrama de dispersão pode ser usado para determinar se existe uma correlação linear (linha reta) entre duas variáveis, ou diversos outros tipos de correlação. Na Figura 3.7, são mostrados 4 diagramas de dispersão. No item (a), o diagrama de dispersão mostra uma correlação linear positiva entre x e y devido a tendência de y aumentar quando o valor de x aumenta. No item (b), há uma correlação linear negativa entre x e y , já que os valores de y tendem a decrescer conforme o valor de x aumenta. Em (c), não é observada uma correlação entre x e y ; em (d), é observada uma correlação entre x e y , porém essa correlação não é linear.

De modo geral, a quantificação do grau de associação entre duas variáveis é feita pelos chamados coeficientes de associação, ou correlação. Esses são medidas que descrevem, por meio de um único número, a associação (ou dependência) entre duas variáveis (MORETTIN; BUSSAB, 2010). A correlação padroniza os valores da covariância fazendo com que o valor nunca fique fora do intervalo $[-1, 1]$. Caso o valor da correlação seja próximo à zero, as duas variáveis não estão relacionadas linearmente, mas podem estar correlacionados de forma não-linear. Já uma correlação positiva indica que as duas variáveis se movem juntas e a relação linear é mais forte quanto mais se aproxima de um. Uma correlação negativa indica que as duas variáveis se movem em direções opostas e a relação linear também fica mais forte quanto mais próxima de -1 .

Enquanto a covariância mede a direção da relação linear entre duas variáveis, a correlação mede tanto a direção da relação linear como a intensidade dessa relação, já que, em uma correlação o valor 1 corresponde a uma relação linear perfeita e como os valores da covariância não são padronizados, é difícil determinar a intensidade da relação entre

as variáveis.

A normalização da covariância pode ser calculada pelo valor da covariância entre as variáveis e divisão desse valor pela multiplicação das variâncias de cada variável. Esse cálculo é dado por

$$\rho_{(y,z)} \equiv \text{corr}(Y, Z) = \sigma_{yz} / \sigma_y \sigma_z = \frac{\sum_{i=1}^n (Y_i - \mu_y)(Z_i - \mu_z)}{\sqrt{\sum_{i=1}^n (Y_i - \mu_y)^2} \sqrt{\sum_{i=1}^n (Z_i - \mu_z)^2}}, \quad (3.1)$$

onde $\text{corr}(Y, Z)$ é a correlação entre as variáveis Y_i e Z_i ; e σ_{yz} é a covariância entre as variáveis Y_i e Z_i . O valor de uma correlação entre duas variáveis é simétrico, sendo assim $\text{corr}(Y, Z) = \text{corr}(Z, Y)$.

Após o cálculo do coeficiente da correlação, o coeficiente deve ser verificado quanto à sua significância. Para tanto, é necessário fazer uma comparação entre o valor absoluto do coeficiente com o valor crítico obtido pelo tamanho da população e pelo nível de significância desejado. Usualmente, os níveis de significância utilizados são 0,01 e 0,05. Quando o nível de significância está configurado em 5% isso quer dizer que é tolerado, em 5% dos casos, que valores de coeficientes não significativos sejam considerados significativos.

Considerando todos os dados da dimensão 1 do vetor de características que descreve todos os pixel como uma variável e todos os dados da dimensão 2 uma outra variável (e assim por diante), para obter o COHAWES da imagem é necessário calcular as correlações entre cada par de variáveis de características construindo uma matriz simétrica $C_{R_{ij}}$ de tamanho 64×64 , para cada região. Nesse cálculo de correlação, é utilizado o valor 0,05 como nível de significância. Esse processo está representado no item (iii) da Fig. 3.1. Cada $C_{R_{ij}}$ denota uma matriz de correlação que é representada por:

$$C_{R_{11}} = \begin{bmatrix} 1 & \rho_{(v_1,v_2)} & \rho_{(v_1,v_3)} & \cdots & \rho_{(v_1,v_{64})} \\ \rho_{(v_2,v_1)} & 1 & \rho_{(v_2,v_3)} & \cdots & \rho_{(v_2,v_{64})} \\ \rho_{(v_3,v_1)} & \rho_{(v_3,v_2)} & 1 & \cdots & \rho_{(v_3,v_{64})} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{(v_{64},v_1)} & \rho_{(v_{64},v_2)} & \rho_{(v_{64},v_3)} & \cdots & 1 \end{bmatrix}, \quad (3.2)$$

onde $C_{R_{11}}$ representa a matriz de correlação da região da linha 1 coluna 1; $\rho_{(v_1,v_2)}$ representa a correlação entre as características v_1 e v_2 , dessa região.

3.4.2 Concatenação do vetor de características

Sendo a matriz de correlação uma matriz simétrica, alguns valores da matriz podem ser desconsiderados na vetorização do COHAWES. Nesse momento, apenas os valores acima da diagonal principal das matrizes de correlação são considerados e para isso é realizado a partir do produto de Hadamard da matriz de correlação com uma matriz triangular superior, onde todos os valores da triangular superior é igual a 1 e todos os valores da diagonal principal é igual a 0. Essa matriz deve ter o mesmo tamanho da matriz de correlação, ou seja 64×64 , e pode ser representado da seguinte forma:

$$C_{R_{ij}} = \begin{bmatrix} 1 & \rho(v_1, v_2) & \rho(v_1, v_3) & \cdots & \rho(v_1, v_{63}) & \rho(v_1, v_{64}) \\ \rho(v_2, v_1) & 1 & \rho(v_2, v_3) & \cdots & \rho(v_2, v_{63}) & \rho(v_2, v_{64}) \\ \rho(v_3, v_1) & \rho(v_3, v_2) & 1 & \cdots & \rho(v_3, v_{63}) & \rho(v_3, v_{64}) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho(v_{63}, v_1) & \rho(v_{63}, v_2) & \rho(v_{63}, v_3) & \cdots & \cdots & \rho(v_{63}, v_{64}) \\ \rho(v_{64}, v_1) & \rho(v_{64}, v_2) & \rho(v_{64}, v_3) & \cdots & \rho(v_{64}, v_{63}) & 1 \end{bmatrix}$$

Figura 3.8: Em cor laranja, valores acima da diagonal principal da matriz de correlação que são considerados durante o processo de vetorização do COHAWES.

$$A = \begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 1 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}. \quad (3.3)$$

Com o produto de Hadamard entre a matriz de correlação e a matriz A , cada valor da matriz de correlação será multiplicado pelo valor que estiver na mesma posição na matriz A . Isso mantém os valores de correlação entre todas as variáveis, porém tornando zero os valores repetidos da matriz simétrica, bem como todos os valores da diagonal principal que, na matriz de correlação, sempre terá 1 como valor dos seus elementos. O produto de Hadamard pode ser expressado da seguinte forma:

$$(C_{R_{11}} \circ A)_{i,j} = (C_{R_{11}})_{i,j} (A)_{i,j}. \quad (3.4)$$

Um exemplo da aplicação do produto de Hadamard pode ser visto na Fig. 3.8, onde a região cinza terá seus valores multiplicados por 0 e a região em laranja terá seus valores mantidos. Após o produto de Hadamard entre $C_{R_{ij}}$ e A apenas os valores diferentes de 0 ($\nabla_{S_{R_{ij}}}$) são vetorizados e a dimensão desse vetor é dada por: $((64 \times 64) - 64)/2$ totalizando 1.984 características. A concatenação de todos os vetores que descrevem cada região (total de 25 regiões) compõe o COHAWES que terá uma dimensão final de 49.600 (25×1.984). Esse processo está representado no item (iv) da Fig. 3.1.

3.5 CONSIDERAÇÕES FINAIS

O cálculo do COHAWES é realizado através da aplicação de filtros horizontais e verticais do tipo Haar wavelet, que permitem uma análise de vizinhança em todos os pixels da imagem. Essa extração se assemelha a extração do descritor do U-SURF citado na Seção 2.2.5. Entretanto, como estamos considerando que todos os pixels da imagem são pontos de interesse, a extração do descritor do pixel é feita de forma densa. Já que todos os pixels da imagem são utilizados para cálculo dos descritores, o detector de pontos de interesse, bem como o cálculo da escala é desprezado no COHAWES. Este descritor aplicado em cada pixel considera uma vizinhança 20×20 permitindo uma descrição discriminativa da textura ao redor do pixel. Além disso, o COHAWES utiliza as matrizes de correlação

que descrevem o nível de relacionamento entre características de textura dentro de sub-regiões da imagem, codificando assim uma descrição que considera na sua estrutura toda a imagem, a textura da vizinhança do pixel, a correlação dentro de sub-regiões e apresenta como resultado valores normalizados entre $[-1 \ 1]$.

AVALIAÇÃO EXPERIMENTAL

4.1	Avaliação experimental do descritor inicial	44
4.1.1	Classificação por meio de métricas de distâncias	45
4.1.2	Classificação supervisionada	47
4.1.3	Seleção de subconjunto de características	48
4.1.4	Utilização de um descritor de textura como características correlacionadas	49
4.2	Avaliando melhor o COHAWES	50
4.2.1	Metodologia	51
4.2.2	Análise comparativa entre COHAWES e outros métodos	52
4.2.3	Análise de resultados	56

A utilização das matrizes de covariância como descritor foi introduzida em (TUZEL; PORIKLI; MEER, 2006), apresentando resultados superiores em avaliações comparativas à época. A partir disso, formulou-se uma hipótese de que o cálculo normalizado do relacionamento entre outras características poderia aumentar o poder discriminativo do descritor. Decidiu-se investigar um grupo de características e novas maneiras de utilizar as regiões locais da imagem para avaliar qual conjunto ou subconjunto de características poderia apresentar maior discriminabilidade para representar faces. Além disso, também uma outra forma de se agregar características com informações de textura foi investigada. Para os dois tipos de avaliações, os data sets FERET (subconjuntos FB, DUP1 e DUP2) (PHILLIPS et al., 1997), IMM (NORDSTRØM et al., 2004), ORL (SAMARIA; HARTER, 1994), Yale (BELHUMEUR; HESPANHA; KRIEGMAN, 1997), extended Yale face database B (GEORGHIADES; BELHUMEUR; KRIEGMAN, 2001) e PF83LFW (BECKER; ORTIZ, 2013) foram utilizados. As principais características destes data sets são sumarizadas na Tabela 4.1.

Tabela 4.1: Resumo dos data sets utilizados nas experimentações.

Dataset	In- divíduos	Total de imagens	Treino	Teste	Características
FERET FA	994	994	-	-	Imagens para treinamento do data set FERET
FERET FB	992	992	FA	FB	Mesmo dia e com a mesma iluminação de FA
FERET DUP1	250	736	FA	DUP1	Após alguns dias que FA
FERET DUP2	75	228	FA	DUP2	Ao decorrer de um ano de FA
IMM	40	240	-	-	Rotação da Face Expressões faciais Variação de iluminação Possível uso de óculos
ORL	40	400	-	-	Momentos diferentes Leve variação de iluminação Expressões faciais Possível uso de óculo
PF83LFW	83	25.068	8.720	4.282	Universo aberto
Yale	15	165	-	-	Momentos diferentes Variação de iluminação Expressões faciais Possível uso de óculos
Yale <i>extended</i> B	38	2.432	-	-	9 poses 64 condições de iluminação

4.1 AVALIAÇÃO EXPERIMENTAL DO DESCRITOR INICIAL

O grupo de características proposto por Tuzel, Porikli e Meer (2006) foi utilizado como base e algumas substituições e acréscimos de características foram feitas, tais como: a remoção do RGB e utilização da escala de cinza da imagem, o acréscimo da magnitude e do gradiente de Sobel e também o acréscimo de dois gradientes intermediários de Sobel (G_{xy} e G_{yx}). Além dessas mudanças, foram acrescentados também os cálculos das características do tipo LBP e *local phase quantization* (LPQ) (OJANSIVU; HEIKKILÄ, 2008), além da aplicação de um filtro de desfocagem (*disk*), com raio 5, e de um filtro laplaciano. Com isso, um conjunto de 15 características foram inicialmente utilizadas, seguindo o seguinte formato de vetor, v_g :

$$v_g = [gray \ x \ y \ G_x \ G_y \ G_{xx} \ G_{xy} \ G_{yx} \ G_{yy} \ G_{mag} \ G_{dir} \ LBP \ LPQ \ disk \ laplacian] , \quad (4.1)$$

onde *gray* é a intensidade do pixel na escala de cinza; x e y são os valores de posicionamento do pixel; G_x e G_y são os gradientes de Sobel na imagem; G_{xx} e G_{xy} são os gradientes de Sobel sobre G_x ; G_{yx} e G_{yy} são os gradientes de Sobel sobre G_y ; G_{mag} é o valor da magnitude do gradiente de Sobel, que é dado por $\sqrt{G_x^2 + G_y^2}$; G_{dir} é a direção do gradiente de Sobel, que é dada por $\arctan(G_x/G_y)$; *LBP* é o cálculo do *local binary patterns* para a imagem; *LPQ* é o cálculo do *local phase quantization* para a imagem; *disk* é a aplicação do filtro de desfocagem com raio 5; e *laplacian* é a aplicação do filtro laplaciano que é dado pela seguinte matriz:

$$\frac{4}{(\alpha + 1)} \begin{bmatrix} \alpha/4 & 1-\alpha/4 & \alpha/4 \\ 1-\alpha/4 & 1 & 1-\alpha/4 \\ \alpha/4 & 1-\alpha/4 & \alpha/4 \end{bmatrix}, \quad (4.2)$$

onde α foi utilizado com o valor de 0.2.

Para a avaliação do vetor, v_g , com as 15 características, foram utilizados data sets de reconhecimento facial, e o seguinte pipeline: os dados acima da diagonal principal da matriz de correlação do vetor que representa a imagem; como classificador métricas de distância; e os data sets IMM (NORDSTRØM et al., 2004), ORL (SAMARIA; HARTER, 1994) e Yale (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) são utilizados para se levantar o desempenho no reconhecimento facial a partir de divisões na imagem e cálculo das correlações entre as características em cada divisão.

4.1.1 Classificação por meio de métricas de distâncias

Foram avaliadas várias métricas de distâncias, porém os melhores resultados foram alcançados com as distâncias euclidiana, Manhattan e distância dos cossenos. As métricas de distância entre os valores das matrizes de correlação são calculadas da seguinte forma: sendo Q a representação dos dados acima da diagonal da matriz de correlação da face-de-busca e T a representação dos dados acima da diagonal da matriz de correlação da face-alvo, a distância euclidiana será dada por:

$$\sqrt{\sum_{i=1}^n (Q_i - T_i)^2}. \quad (4.3)$$

A distância de Manhattan é uma métrica da geometria euclidiana onde a distância entre dois pontos é dada pela soma das diferenças absolutas de suas coordenadas e pode ser expressada da seguinte forma:

$$\sum_{k=1}^n |T_k - Q_k|. \quad (4.4)$$

Outra métrica de distância avaliada foi a distância dos cossenos, que é um termo usado frequentemente para o complemento no espaço positivo da semelhança do cosseno e é expressada por:

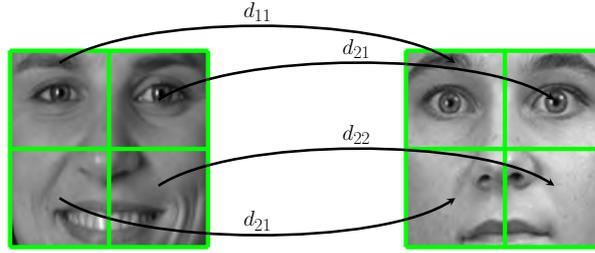


Figura 4.1: Acúmulo das distâncias entre os segmentos das imagens.

$$d_{QT} = 1 - \frac{\sum_{i=1}^n Q_i T_i}{\sqrt{\sum_{i=1}^n Q_i^2} \sqrt{\sum_{i=1}^n T_i^2}}. \quad (4.5)$$

Além de diferentes métricas de distâncias, a divisão da imagem em regiões imagéticas (*patches*) também foi avaliada. A imagem passou a ser dividida em *patches* e para cada *patch* era calculada uma matriz de correlação. Nesse cenário, a distância final entre um par de imagens é dada pelo somatório das distâncias das matrizes de correlação de *patches* equivalentes. A Figura 4.1 ilustra a equivalência dos *patches*, e a distância final entre cada par de imagens passou a ser dada por:

$$D = \sum_{i=1}^n \sum_{j=1}^n d_{ij}, \quad (4.6)$$

onde d_{ij} é a distância entre o *patch* ij da face-de-busca e o *patch* ij da face-alvo. Com essa abordagem foram definidos 6 modos de dividir a imagem em *patches* que são eles: 2×2 , 4×4 , 5×5 , 7×7 , 8×8 e 10×10 .

Como os data sets IMM, ORL e Yale não possuem um protocolo de reconhecimento estabelecido, foram escolhidas, de forma aleatória, 75% das imagens de cada pessoa do data set para face-alvo e 25% para face-de-busca. Os resultados alcançados para esta avaliação são sumarizados na Tabela 4.2, onde é mostrada a precisão de acerto de reconhecimento, utilizando as três distâncias e as diversas formas de se dividir a imagem.

Como a utilização do descritor com a correlação das 15 características definidas na Equação 4.1 e das métricas de distância euclidiana, Manhattan e distância do cosseno a precisão do reconhecimento facial não apresentou valores próximos aos valores da reprodução do trabalho de (DING et al., 2016) utilizando DCP e esses valores podem ser visto na Tabela 4.3.

Utilizando métricas de distâncias como classificação não supervisionada a média do percentual de precisão no reconhecimento das faces-de-busca dos data sets não apresentaram mudanças significativas. Para a métrica de distância euclidiana a média de reconhecimento foi de 56,51%, para distância dos cossenos foi de 57,46% e a distância

Tabela 4.2: Precisão do reconhecimento de face para distância euclidiana, Manhattan e distância do cosseno com o vetor de 15 características v_g .

Distância	Data sets	Divisões da imagem						
		Holística	2×2	4×4	5×5	7×7	8×8	10×10
Euclidiana	IMM	13,33%	17,50%	22,92%	22,92%	26,67%	28,75%	28,33%
	ORL	66,75%	76,75%	81,25%	79,50%	77,75%	78,75%	78,25%
	Yale	59,39%	61,21%	67,88%	70,30%	75,15%	75,15%	78,18%
Manhattan	IMM	15,42%	18,75%	25,83%	25,00%	29,17%	29,17%	30,00%
	ORL	70,50%	77,25%	81,50%	79,25%	78,50%	78,50%	78,75%
	Yale	63,64%	64,24%	69,09%	72,12%	75,15%	75,15%	80,00%
Cosseno	IMM	14,17%	16,67%	19,17%	20,00%	25,42%	25,42%	27,08%
	ORL	73,00%	80,50%	84,00%	81,50%	80,25%	80,00%	80,00%
	Yale	62,42%	63,64%	71,52%	74,55%	73,94%	75,76%	77,58%

de Manhattan apresentou a maior média de 57,95%. Uma diferença de apenas 1,44%. Devido à pouca variação na precisão do reconhecimento das faces-de-busca com as mudanças nas métrica de distância criou-se a hipótese de que um método de classificação supervisionado poderia aumentar o percentual de acertos.

4.1.2 Classificação supervisionada

Considerando o pouco impacto das mudanças nas métricas de distância decidiu-se experimentar uma SVM para ser utilizada como classificador no reconhecimento facial com o descritor proposto. Os testes foram realizados nos data sets IMM, ORL, e Yale e podem ser vistos da Tabela 4.4.

Com a utilização da SVM como método de classificação supervisionada a média do percentual de precisão no reconhecimento considerando os data sets IMM, ORL e Yale passou a ser 77,91%, apresentando um aumento de 19,96% em relação à melhor média utilizando métricas de distâncias. Com esse experimento a hipótese do classificador supervisionado se mostrou verdadeira e essa estratégia passou a ser adotada para os próximos testes. Apesar da melhoria na precisão do acerto no reconhecimento facial os percentuais ainda estão abaixo dos valores da tabela 4.3.

Tabela 4.3: Precisão do reconhecimento de face utilizando DCP. Dados da reprodução do trabalho (DING et al., 2016).

Data sets	Precisão
IMM	72,5%
ORL	100,0%
Yale	100,0%

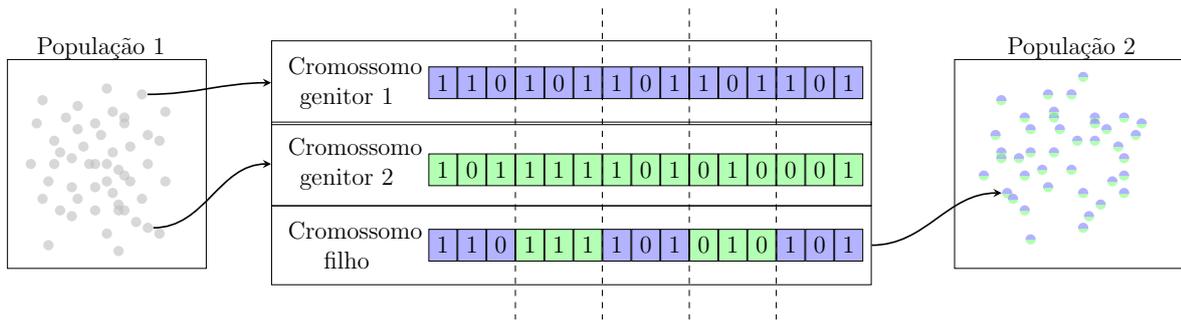


Figura 4.2: Utilização de um método de busca baseado em algoritmo genético na seleção do conjunto de características para representar as imagens de faces.

4.1.3 Seleção de subconjunto de características

Como a discriminabilidade da representação proposta para a imagem não era suficiente para realizar a identificação da face-de-busca com valores de precisão superiores ao do DCP, valores da tabela 4.3, formulou-se uma hipótese de que mudanças no grupo de características do vetor de representação poderia aumentar o poder discriminativo do descritor. Decidiu-se investigar por subconjuntos das 15 características que apresentasse uma maior discriminabilidade para a representação de faces.

Foi utilizado um método de busca baseado em algoritmo genético para encontrar o subconjunto das 15 características que maximizasse o percentual de acertos no reconhecimento da face-de-busca. O algoritmo genético estabelece uma população inicial de possíveis soluções e seleciona aleatoriamente indivíduos dessa população para avaliar por uma função objetivo. A função objetivo indica o quão próximo da solução ótima aquele indivíduo está. Em cada etapa, os indivíduos avaliados que apresentaram melhores resultados são selecionados e utilizados para gerar uma nova população de possíveis soluções. A utilização dos indivíduos para gerar uma nova população é inspirado no processo de recombinação genética onde, cada indivíduo é visto como um cromossomo e cruzamentos entre os cromossomos, ou mutações, geram novos cromossomos. Ao longo de sucessivas gerações, a população “evolui” para encontrar uma solução ótima.

Na busca do subconjunto de características “ótimo” foi utilizado como cromossomo

Tabela 4.4: Percentuais de acerto de reconhecimento de identidade usando SVM como classificador e matriz de correlação do vetor v_g de 15 características como descritor.

Data sets	<i>support vector machine</i>						
	Holística	2×2	4×4	5×5	7×7	8×8	10×10
IMM	25,00%	49,38%	61,88%	58,75%	58,13%	56,88%	54,38%
ORL	80,00%	90,71%	92,50%	91,79%	92,50%	90,71%	91,07%
Yale	73,33%	82,50%	95,00%	96,67%	98,33%	98,33%	98,33%

Tabela 4.5: Percentuais de acerto de reconhecimento de identidade utilizando o grupo de 5 característica selecionadas pelo algoritmo genético (v_{ag}).

Data sets	<i>support vector machine</i>					
	2×2	4×4	5×5	7×7	8×8	10×10
IMM	25,63%	51,25%	48,75%	57,50%	53,75%	56,88%
ORL	61,43%	82,50%	85,36%	90,00%	88,93%	88,21%
Yale	75,83%	83,33%	93,33%	95,00%	95,83%	97,50%

uma palavra de 15 bits onde, cada bit representa a utilização de uma das 15 características. O bit com valor 1 significa que a característica que esteja nessa mesma posição, definida na Equação 4.1, fará parte do vetor de representação da imagem, o bit com valor 0 significa que a característica não fará parte. A Figura 4.2 ilustra a utilização do algoritmo genético para a escolha do conjunto de características para ser utilizado no reconhecimento da face-de-busca. A execução do algoritmo genético utilizou como função objetivo maximizar a taxa de acerto no reconhecimento facial, utilizando como data set o FERET. Esse data set foi escolhido por ser o único dos listados na Tabela 4.1 que possui um protocolo de reconhecimento estabelecido e nesse protocolo apenas uma única imagem é utilizada como face-alvo, tornando o data set um dos mais desafiadores. O algoritmo genético encontrou o cromossomo $[1\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 1]$ que representa o vetor com as seguintes características:

$$v_{ag} = [gray\ G_x\ G_{mag}\ LPQ\ laplacian] . \quad (4.7)$$

Esse novo grupo de características (chamado v_{ag}) passou a ser utilizado para representar as imagens de face e com ele foram realizados os testes de precisão de reconhecimento da imagem-de-busca nos data sets ORL, IMM e Yale. Na Tabela 4.5 são mostrados os valores da precisão do reconhecimento utilizando o vetor v_{ag} com as imagens divididas em 2×2 , 4×4 , 5×5 , 7×7 , 8×8 e 10×10 . O cálculo da matriz de correlação utilizando o vetor de representação de forma holística deixou de ser avaliado por ter apresentado sempre os percentuais mais baixos.

A utilização do v_{ag} como representação da imagem, montado através do cromossomo escolhido pelo algoritmo genético para o data set FERET, apresentou resultados piores do que utilizando o vetor com todas as 15 características (v_g). Com isso, foi elaborada uma nova hipótese de que um descritor multidimensional que avaliassem a vizinhança dos pixels de forma robusta poderia substituir o grupo das características utilizados para compor o vetor de representação das imagens de face.

4.1.4 Utilização de um descritor de textura como características correlacionadas

A hipótese de utilizarmos um descritor de textura para, a partir das correlações, tentar obter um descritor mais discriminativo ganhou força, após o fraco desempenho do grupo e subgrupos com as 15 características iniciais. Nesse ponto, optou-se por explorar o descritor utilizado no SURF(BAY; TUYTELAARS; GOOL, 2006), o qual descreve cada

Tabela 4.6: Comparação do descritor utilizando matriz de correlação dos somatórios de gradientes de filtros do tipo Haar wavelet (COHAWES) e divisão 5×5 com correlação do vetor com 15 característica (v_g) e divisão 5×5 , correlação do vetor com as característica selecionadas pelo algoritmo genético (v_{ga}) e divisão 5×5 e DCP (DING et al., 2016). Classificador implementação SVM de (ORTIZ; BECKER, 2014).

Descritor	Data sets			
	IMM	ORL	Yale	Yale <i>extended</i> B
COHAWES	85.0%	100.0%	100.0%	99.7%
v_g	72.5%	97.5%	100.0%	99.8%
v_{ga}	75.0%	97.5%	93.3%	96.3%
DCP	72.5%	100.0%	100.0%	99.8%

pixel com somatórios de gradientes de filtros do tipo Haar wavelet em uma vizinhança de 20×20 , construindo um vetor de tamanho 64 para descrever um único pixel.

Um novo vetor de representação das imagens de face foi construído com os somatórios dos gradientes da aplicação de dois filtros do tipo Haar wavelet, um vertical e outro horizontal. Com esse novo vetor foram realizados os testes de precisão de reconhecimento da imagem-de-busca nos data sets ORL, IMM, Yale e Yale *extended* B (GEORGHIADIS; BELHUMEUR; KRIEGMAN, 2001). Na Tabela 4.6 são sumarizados os resultados da precisão do reconhecimento utilizando correlação do vetor com 15 características (v_g), correlação do vetor com as características selecionadas pelo algoritmo genético (v_{ag}) e correlação do novo vetor com os somatórios dos gradientes de Haar wavelet (COHAWES), com a divisão em 5×5 da imagem. A escolha de 5×5 foi devido ao fato dessa divisão apresentar valores de precisão de reconhecimento próximos da média em todos os data sets. A média de precisão do COHAWES foi de 96,18%, sendo essa a maior entre todos os vetores, o vetor com as 15 características iniciais v_g teve uma média de 92,45%, o v_{ag} teve 89,90% e o COHAWES obteve uma média maior do que o DCP que teve 93,07% de média. Com esse experimento a hipótese do descritor multidimensional para substituir as 15 características iniciais foi considerada verdadeira fazendo com que os somatórios dos gradientes da aplicação de dois filtros do tipo Haar wavelet fosse explorado em outros experimentos.

4.2 AVALIANDO MELHOR O COHAWES

A avaliação experimental do COHAWES foi realizada com o auxílio de um *framework* que possibilitou a parametrização de pontos da execução do reconhecimento facial, como por exemplo: grupo de características; tipo de relação entre as características; quantidade e tamanho das regiões que a imagem é dividida; método de comparação entre a face-de-busca e face-alvo; e data set utilizado para verificar o reconhecimento. A parametrização desses pontos permitiu configurar, e executar, vários cenários para avaliar a precisão do reconhecimento facial. A Figura 4.3 apresenta as etapas desse processo.

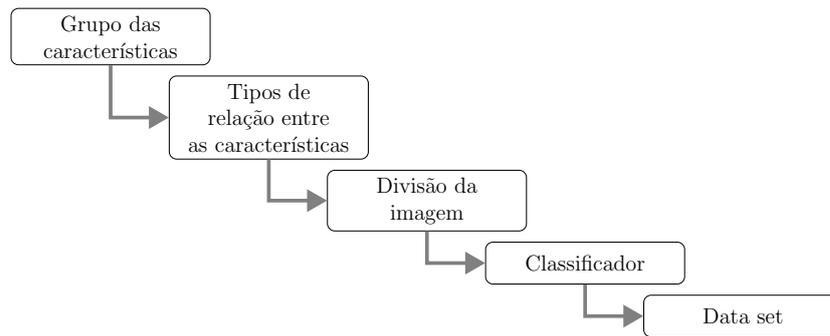


Figura 4.3: Etapas gerais do *framework* que permite a escolha do grupo das características, tipos de relação entre essas características, forma de divisão da imagem, classificador e data set.

4.2.1 Metodologia

A metodologia utilizada para avaliar a discriminabilidade do COHAWES foi através de comparação da precisão de reconhecimento facial em cenários com descritores, classificadores e data sets diferentes. Inspirado nos testes realizados em (ORTIZ; BECKER, 2014), foram definidos 9 classificadores, 8 data sets e 7 descritores para compor os cenários de testes para avaliar o COHAWES. A Figura 4.4 ilustra os 504 cenários possíveis.

Os classificadores utilizados nos testes para avaliar o COHAWES foram: (a) *Nearest Neighbor* (NN) (COVER; HART, 1967); (b) *Locally-constrained Linear Coding* (LLC) (WANG et al., 2010); (c) *K Nearest Neighbor com Sparse Representation Clas-*

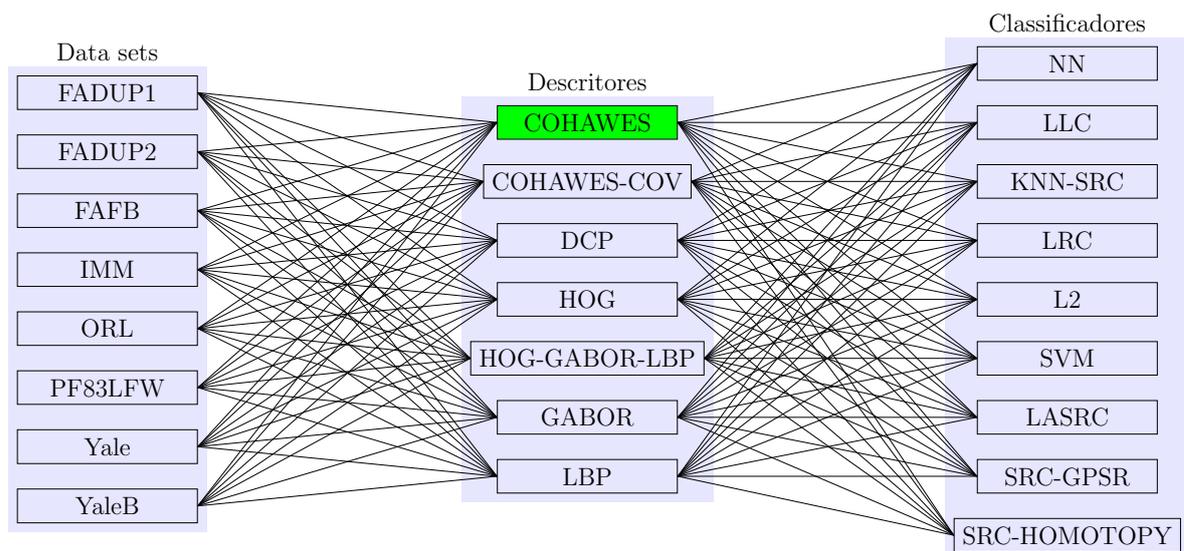


Figura 4.4: Cenários de testes considerando 8 data sets, 7 descritores e 9 classificadores totalizando 504 cenários.

sifier(KNN-SRC) (ZHANG; YANG, 2010); (d) *Linear Regression Classification* (LRC) (NASEEM; TOGNERI; BENNAMOUN, 2010); (e) Minimização ortonormal ℓ_2 (MO-L2) (SHI et al., 2011); (f) *Support-Vector Machines* (SVM) (HEISELE; HO; POGGIO, 2001); (g) *Linearly Approximated Sparse Representation-based Classification* (LASRC) (ORTIZ; BECKER, 2014); (h) *Sparse Representation Classifier com Gradient Projection for Sparse Reconstruction* (SRC-GPSR) (FIGUEIREDO; NOWAK; WRIGHT, 2007); (i) *Sparse Representation Classifier com Homotopy*(MALIOUTOV; CETIN; WILLISKY, 2007) (SRC-Homotopy).

4.2.2 Análise comparativa entre COHAWES e outros métodos

Em (ORTIZ; BECKER, 2014), os classificadores são testados com os descritores Gabor, HOG e LBP, e com um descritor que associa esses três. Foi acrescentado à essa relação de descritores o DCP, proposto por Ding et al. (2016), e uma variação do COHAWES que é calculada com a matriz de covariância (COHAWES-COV) com a intenção de avaliar a influência da correlação na discriminabilidade do descritor.

Todos os resultados apresentados nas Tabelas 4.7 e 4.8 são os valores percentuais da precisão de acerto do reconhecimento facial considerando os 3 itens para compor o cenário: o descritor, o classificador e o data set. Nessas duas tabelas, estão destacados os valores das maiores precisões de reconhecimento facial obtidos por um descritor para um classificador (colunas) naquele data set. Nos casos de acontecer um empate no maior valor de precisão de reconhecimento facial todos os descritores que tenham o valor de precisão igual ao máximo, naquele classificador, terão destaques em negrito na tabela.

Nos cenários apresentados na Fig. 4.4, o COHAWES e o COHAWES-COV são comparados com os seguinte descritores: (a) DCP; (b) HOG; (c) Um vetor combinando HOG, Gabor wavelets e LBP (HOG+GABOR+LBP); (d) Gabor wavelets (GABOR); (e) LBP. Vale destacar que os data sets do FERET possuem uma separação entre as face-de-busca das imagens face-alvo. O subconjunto de imagens chamadas de FA são as imagens que são utilizadas para o treino dos classificadores (face-alvo) já as imagens dos subconjuntos FB, DUP1 e DUP2 são utilizadas como face-de-busca seguindo o protocolo definido em (PHILLIPS et al., 2000). Os outros data sets, que não possuem um protocolo definido separando o treino do teste, foram utilizadas 75% das imagens para treino e 25% para teste. Na tabela 4.1 pode ser visto um resumo dos data sets utilizados nas experimentações.

Para facilitar a comparação dos resultados do COHAWES com os resultados dos outros 6 descritores em cada classificador os testes realizados com o data set FERET são apresentados em forma de gráfico na Fig 4.5. O valor da precisão do COHAWES é representado pela área em azul e o valor dos outros descritores são representados por barras. Os valores da precisão em reconhecimento facial estão agrupados por classificador e esses classificadores estão identificados na parte inferior do gráfico. Essa organização do gráfico permite que seja feita uma comparação visual, em cada classificador, do valor da precisão do COHAWES com todos os outros descritores ao mesmo tempo. No dataset FERET o DCP se mostrou mais discriminativo do que os outros descritores e o COHAWES ficou como a segunda maior precisão. Na Tabela 4.7 são mostrados os valores dos percentuais de precisão no FERET.

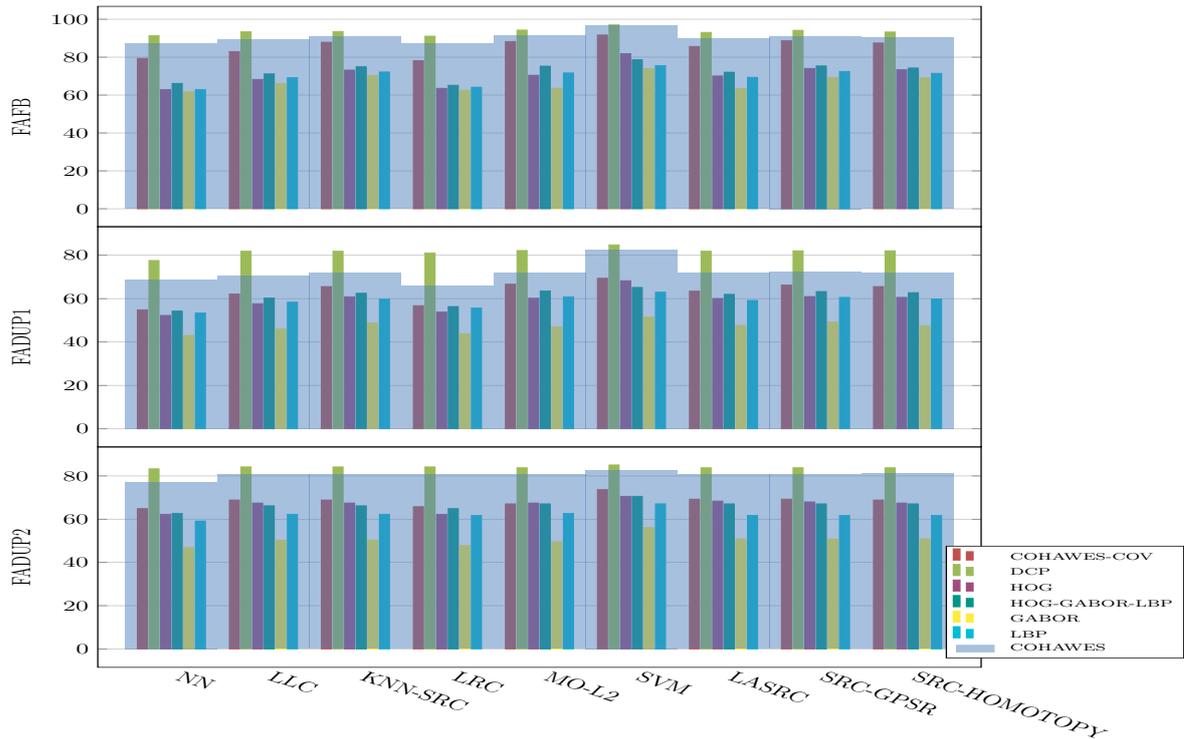


Figura 4.5: Comparação dos resultados do COHAWES com os outros descritores no data set FERET: FAFB, FADUP1 e FADUP2.

Tabela 4.7: Precisão de reconhecimento facial de todos os descritores e todo os classificadores no FERET.

DATA SET	Descritor	NN	LLC	KNN-SRC	LRC	MO-L2	SVM	LASRC	SRC-GPSR	SRC-Homotopy
FADUP1	COHAWES	68.5%	70.4%	71.7%	65.9%	71.7%	82.5%	71.9%	72.3%	72.0%
	COHAWES-COV	54.9%	62.2%	65.6%	56.8%	66.7%	69.4%	63.5%	66.3%	65.6%
	DCP	77.6%	81.9%	81.9%	81.1%	82.2%	84.8%	81.9%	82.1%	82.1%
	HOG	52.3%	57.7%	60.9%	53.9%	60.2%	68.3%	60.1%	61.0%	60.6%
	HOG+GABOR+LBP	54.3%	60.3%	62.5%	56.4%	63.6%	65.2%	62.0%	63.3%	62.8%
	GABOR	43.1%	46.1%	48.8%	43.9%	47.1%	51.5%	47.7%	49.2%	47.6%
	LBP	53.4%	58.4%	59.8%	55.7%	60.9%	63.0%	59.2%	60.6%	59.9%
FADUP2	COHAWES	77.2%	80.7%	80.7%	80.7%	80.7%	82.9%	80.7%	80.7%	81.1%
	COHAWES-COV	64.9%	68.9%	68.9%	65.8%	67.1%	73.7%	69.3%	69.3%	68.9%
	DCP	83.3%	84.2%	84.2%	84.2%	83.8%	85.1%	83.8%	83.8%	83.8%
	HOG	62.3%	67.5%	67.5%	62.3%	67.5%	70.6%	68.4%	68.0%	67.5%
	HOG+GABOR+LBP	62.7%	66.2%	66.2%	64.9%	67.1%	70.6%	67.1%	67.1%	67.1%
	GABOR	46.9%	50.4%	50.4%	47.8%	49.6%	56.1%	50.9%	50.9%	50.9%
	LBP	59.2%	62.3%	62.3%	61.8%	62.7%	67.1%	61.8%	61.8%	61.8%
FAFB	COHAWES	87.3%	89.2%	90.6%	87.3%	91.3%	96.7%	89.9%	91.1%	90.4%
	COHAWES-COV	79.4%	83.0%	87.9%	78.2%	88.3%	91.7%	85.7%	88.7%	87.5%
	DCP	91.4%	93.4%	93.6%	91.1%	94.3%	97.1%	93.1%	94.2%	93.3%
	HOG	62.9%	68.3%	73.2%	63.6%	70.5%	81.9%	70.1%	74.1%	73.4%
	HOG+GABOR+LBP	66.2%	71.2%	75.0%	65.2%	75.3%	78.7%	72.1%	75.4%	74.3%
	GABOR	61.7%	66.1%	70.3%	62.6%	63.7%	73.9%	63.4%	69.3%	69.1%
	LBP	62.9%	69.2%	72.2%	64.2%	71.7%	75.6%	69.4%	72.5%	71.5%

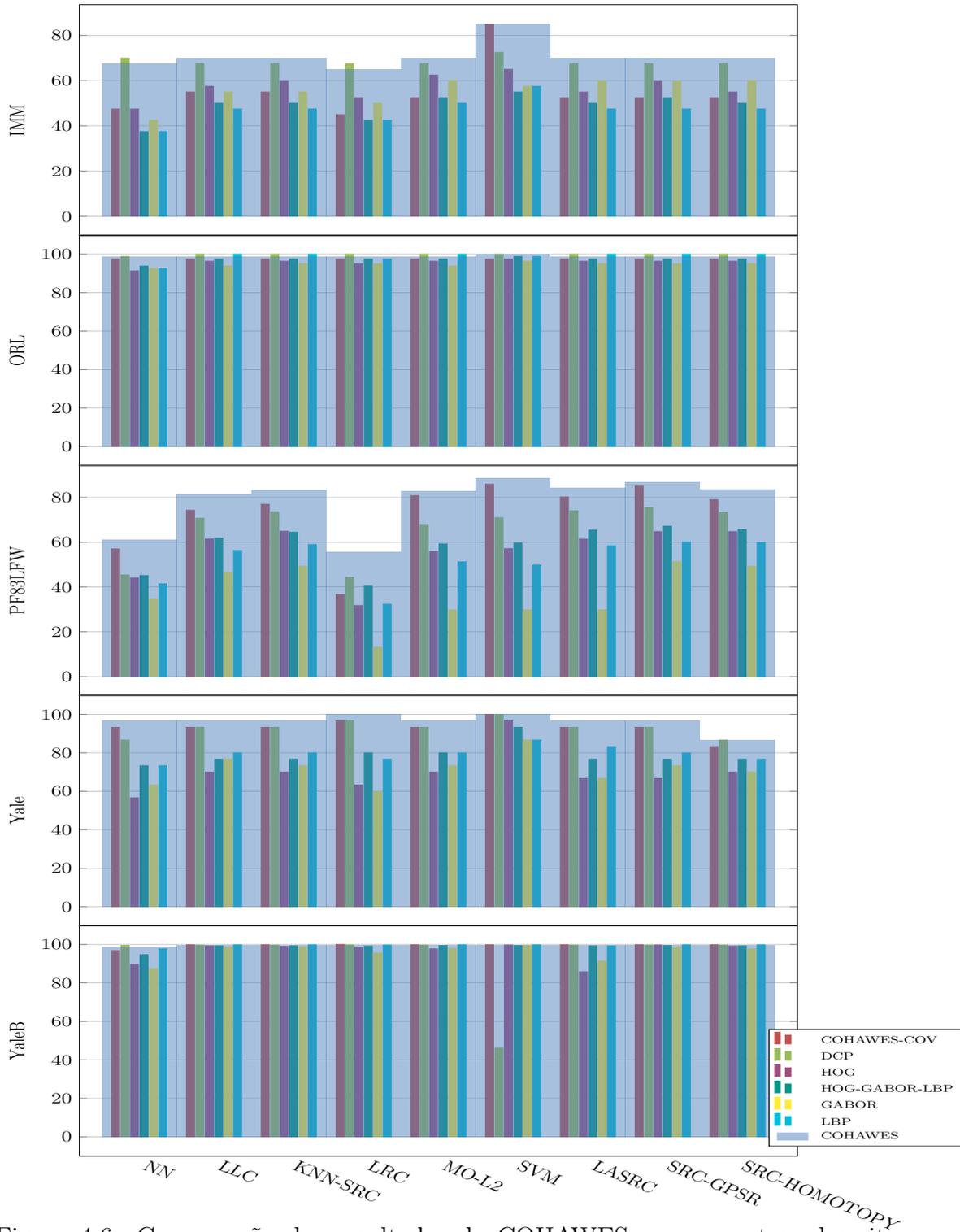


Figura 4.6: Comparação dos resultados do COHAWES com os outros descritores nos data sets IMM, ORL, PF83LFW, Yale e YaleB.

Na Tabela 4.9 são apresentados dois valores de média de precisão em reconhecimento

Tabela 4.8: Precisão de reconhecimento facial de todos os descritores e todos os classificadores nos data sets IMM, ORL, PF83LFW, Yale e Yale *extended* B.

DATA SET	Descritor	NN	LLC	KNN-SRC	LRC	MO-L2	SVM	LASRC	SRC-GPSR	SRC-Homotopy
IMM	COHAWES	67.5%	70.0%	70.0%	65.0%	70.0%	85.0%	70.0%	70.0%	70.0%
	COHAWES-COV	47.5%	55.0%	55.0%	45.0%	52.5%	85.0%	52.5%	52.5%	52.5%
	DCP	70.0%	67.5%	67.5%	67.5%	67.5%	72.5%	67.5%	67.5%	67.5%
	HOG	47.5%	57.5%	60.0%	52.5%	62.5%	65.0%	55.0%	60.0%	55.0%
	HOG+GABOR+LBP	37.5%	50.0%	50.0%	42.5%	52.5%	55.0%	50.0%	52.5%	50.0%
	GABOR	42.5%	55.0%	55.0%	50.0%	60.0%	57.5%	60.0%	60.0%	60.0%
	LBP	37.5%	47.5%	47.5%	42.5%	50.0%	57.5%	47.5%	47.5%	47.5%
ORL	COHAWES	98.8%	98.8%	98.8%	98.8%	98.8%	100.0%	98.8%	98.8%	98.8%
	COHAWES-COV	97.5%	97.5%	97.5%	97.5%	97.5%	97.5%	97.5%	97.5%	97.5%
	DCP	98.8%	100.0%							
	HOG	91.3%	96.3%	96.3%	95.0%	96.3%	97.5%	96.3%	96.3%	96.3%
	HOG+GABOR+LBP	93.8%	97.5%	97.5%	97.5%	97.5%	98.8%	97.5%	97.5%	97.5%
	GABOR	92.5%	93.8%	95.0%	95.0%	93.8%	96.3%	95.0%	95.0%	95.0%
	LBP	92.5%	100.0%	100.0%	97.5%	100.0%	98.8%	100.0%	100.0%	100.0%
PF83LFW	COHAWES	61.2%	81.4%	83.0%	55.5%	82.9%	88.4%	84.1%	86.6%	83.3%
	COHAWES-COV	57.0%	74.3%	76.9%	36.7%	80.8%	85.9%	80.3%	85.1%	79.0%
	DCP	45.4%	70.7%	73.6%	44.4%	68.0%	71.0%	74.1%	75.5%	73.3%
	HOG	44.1%	61.5%	64.9%	31.8%	55.9%	57.2%	61.4%	64.8%	64.8%
	HOG+GABOR+LBP	45.2%	61.9%	64.5%	40.8%	59.3%	59.7%	65.5%	67.2%	65.7%
	GABOR	34.8%	46.4%	49.3%	13.1%	29.8%	29.8%	29.9%	51.4%	49.3%
	LBP	41.5%	56.4%	59.0%	32.3%	51.3%	49.8%	58.4%	60.1%	60.0%
Yale	COHAWES	96.7%	96.7%	96.7%	100.0%	96.7%	100.0%	96.7%	96.7%	86.7%
	COHAWES-COV	93.3%	93.3%	93.3%	96.7%	93.3%	100.0%	93.3%	93.3%	83.3%
	DCP	86.7%	93.3%	93.3%	96.7%	93.3%	100.0%	93.3%	93.3%	86.7%
	HOG	56.7%	70.0%	70.0%	63.3%	70.0%	96.7%	66.7%	66.7%	70.0%
	HOG+GABOR+LBP	73.3%	76.7%	76.7%	80.0%	80.0%	93.3%	76.7%	76.7%	76.7%
	GABOR	63.3%	76.7%	73.3%	60.0%	73.3%	86.7%	66.7%	73.3%	70.0%
	LBP	73.3%	80.0%	80.0%	76.7%	80.0%	86.7%	83.3%	80.0%	76.7%
Yale Ex B	COHAWES	99.0%	99.7%	99.7%	99.8%	99.7%	99.7%	99.7%	99.7%	99.7%
	COHAWES-COV	96.8%	99.8%	99.8%	100.0%	99.8%	99.8%	99.8%	99.8%	99.8%
	DCP	99.5%	99.7%	99.7%	99.8%	99.7%	46.2%	99.7%	99.7%	99.7%
	HOG	89.7%	99.3%	99.0%	98.5%	97.7%	99.7%	85.8%	99.7%	99.2%
	HOG+GABOR+LBP	94.7%	99.3%	99.3%	99.2%	99.5%	99.5%	99.3%	99.5%	99.3%
	GABOR	87.5%	98.5%	98.7%	95.5%	98.0%	99.5%	91.3%	98.7%	97.8%
	LBP	97.7%	99.8%	99.8%	99.7%	99.8%	99.8%	99.3%	99.8%	99.0%

de face para cada descritor. No primeiro valor são considerados todos os classificadores listados em Figura 4.4 e todos os data sets, no segundo é apresentada a média considerando apenas o valor do SVM. A escolha do SVM é por esse classificador ter sido o que apresentou o maior média de precisão levando em conta todos os descritores. Essa Tabela 4.9 mostra, em destaque, que o COHAWES obteve os melhores resultados nas duas médias.

Tabela 4.9: Média dos percentuais de acerto para cada descritor.

Descritor	Média Geral	Média com SVM
COHAWES	85,89%	91,90%
COHAWES-COV	79,51%	87,88%
DCP	85,05%	82,09%
HOG	71,62%	79,61%
HOG+GABOR+LBP	72,88%	77,60%
GABOR	64,90%	68,91%
LBP	71,05%	74,79%

4.2.3 Análise de resultados

A separação de espaços obtida com os classificadores supervisionados ofereceu uma grande diferença nas avaliações da discriminabilidade do descritor proposto. Essa diferença foi de aproximadamente 20 pontos percentuais maior na média da precisão do reconhecimento em comparação com os mesmos dados quando eram avaliados com métricas de distância e analisando as Tabelas 4.7 e 4.8 percebe-se que o SVM foi o classificador que apresentou, em médio, os melhores resultados com todos os descritores. A discriminabilidade do descritor está relacionada também com a escolha do grupo de características que será utilizado no cálculo da relação estatística, seja essa relação covariância ou correlação, mas essa discriminabilidade não está diretamente relacionada com o tamanho do grupo de característica. A partir de um método de busca baseado em algoritmo genético, observou-se que subconjuntos desse grupo de características podem ser selecionados para maximizar a precisão no reconhecimento em data sets específicos, mas que isso não garante melhores resultados em outros data sets.

O descritor COHAWES, que considera uma relação de interdependência entre suas características, apresentou melhor resultado nas avaliações realizadas e foi o descritor que obteve a maior média de reconhecimento facial. Apesar de no data set Yale *extended* B, a covariância das características do COHAWES (a variação COHAWES-COV) ter uma melhor precisão no reconhecimento facial do que a correlação, no geral, a correlação apresentou melhores resultados do que a covariância. Isso pode ser verificado na Tabela 4.9 que mostra as médias das precisões e posiciona o COHAWES como o primeiro com uma precisão geral de 85,89% e uma precisão com SVM de 91,90% já o COHAWES-COV ficou em terceiro com uma precisão geral de 79,51% e uma precisão com SVM de 87,88%.

Vale ressaltar que no data set PF83LFW, ocorreu a maior diferença de percentual de precisão entre o COHAWES e os outros descritores. Naquele data set, o COHAWES apresentou um percentual de precisão 12% superior, em média, em todos os classificadores, em comparação ao DCP que ficou com o terceiro melhor percentual de precisão. Na Tabela 4.1, podemos verificar que esse é o único data set, dentre os utilizados nos testes, que tem como característica ser “universo aberto”, onde não existe controle algum sobre as condições em que a foto é tirada e normalmente é composto por fotos publicadas na

internet. Ainda no data set PF83LFW o COHAWES-COV ficou com o segundo melhor percentual de precisão apresentando 6% superior, em média, em todos os classificadores, em comparação ao DCP. Isso indica que descritores que consideram uma relação de interdependência entre suas características podem ser mais robustos para data sets do tipo “universo aberto”.

DISCUSSÃO E CONCLUSÕES

5.1	Data set de “universo aberto”	59
5.2	Tempo para computar do COHAWES	60
5.3	Significância do coeficiente de correlação	60
5.4	Dispersão de dados diferentes com mesma correlação	61
5.5	Benchmark para reconhecimento facial	61

Os resultados apresentados neste trabalho evidenciam que o descritor COHAWES apresenta precisão de reconhecimento facial comparáveis, e em alguns casos superior, à descritores recentes como o DCP proposto por Ding et al. (2016). Foram realizados testes em data sets com características variadas como: intensa variação na iluminação das faces dos data sets Yale; mudanças faciais decorridas pela passagem do tempo do data set FERET; mudanças nas expressões faciais em todos os data sets; possíveis utilização de óculos nos data sets FERET e ORL; rotação na face no data set IMM; e “universo aberto” do data set PF83LFW.

5.1 DATA SET DE “UNIVERSO ABERTO”

O conceito de data sets de “universo aberto” foi projetado para estudar o problema do reconhecimento de face sem restrições na captura das imagens. O data set *labeled faces in the wild* (LFW) (HUANG et al., 2007) foi o pioneiro nesse conceito e construiu um conjunto de mais de 13.000 imagens de faces coletadas na web de 1.680 das pessoas diferentes. Nesse data set são encontradas imagens com grande variação na pose, iluminação, expressão facial, ambiente, tipo de câmera, parâmetros condições para aquisição da imagem, etc, e essa variedade de situações possibilita que o sistema de reconhecimento facial seja testado em condições mais próximas de situações práticas. Outro data set que seguiu a mesma linha do LFW é o *public figures face database* (PubFig) (KUMAR et al., 2009) de imagens de figuras públicas (celebridades e políticos) adquiridas da internet sujeito

às mesmas variedades de situações apresentadas no LFW. Consequentemente o data set PF83LFW, por ser uma junção de subconjuntos de imagens dos data sets LFW e PubFig (considerando apenas 83 indivíduos) apresenta também uma grade variedade de situações que o descritor de um sistema de reconhecimento facial deva conseguir resolver. No site do PF83LFW¹ existe uma área de resultados reportados onde o descritor que associa Gabor, HOG e LBP em conjunto com o classificador LASRC apresenta o maior valor de precisão de reconhecimento 84,4% mas o COHAWES com SVM, em nossos testes apresentou um valor de precisão de reconhecimento de 88,4%.

5.2 TEMPO PARA COMPUTAR DO COHAWES

Não foi considerado como escopo desse trabalho um estudo sobre o tempo de processamento para extração e comparação do COHAWES porém, os estudos de Bay, Tuytelaars e Gool (2006) mostram que o uso da integral da imagem, que se tornou popular no trabalho de Viola e Jones (2004), reduz drasticamente o tempo de computação para o cálculo dos gradientes dos filtro do tipo Haar wavelet. No trabalho de Bay, Tuytelaars e Gool (2006) o SURF é apresentado como um detector-descritor que gasta 354 milissegundos para ser computado. O COHAWES também calcula os somatórios dos gradientes de Haar wavelet mas não realiza a etapa de detecção, em contra partida realiza o cálculo da matriz de correlação. Apesar de serem calculadas uma matriz para cada *patch* esse cálculo pode ser paralelizado já que não existe dependência entre os valores das matrizes de *patches* diferentes. Isso sugere que, como trabalho futuro, uma implementação do COHAWES com computação paralela pode ser criada e o tempo de extração do descritor deve ser mensurado.

5.3 SIGNIFICÂNCIA DO COEFICIENTE DE CORRELAÇÃO

Em estatística, aborda-se a questão da significância de um resultado usando-se o conceito de hipótese nula. A hipótese nula simplesmente assume que o dado de uma correlação foi obtido apenas por acaso, devido a flutuações probabilísticas dos eventos que estão sendo medidos, e não devido a um efeito real que cause o resultado. Sempre que se trabalha com uma hipótese para explicar um dado fenômeno, deve ser considerada a possibilidade de pelo menos uma hipótese concorrente a ela. No caso da estatística, a hipótese concorrente é chamada de hipótese alternativa e nessa hipótese considera que o resultado da correlação foi obtido devido a uma real relação entre os eventos medidos e que a correlação é significativa. Considerando que hipótese alternativa esteja correta deve ser verificado a probabilidade de que ela ocorra. E quanto menor a for a quantidade de amostras para o cálculo da correlação maior será a probabilidade do valor do coeficiente ter sido obtido pelo acaso. Considerando isso, o tamanho das regiões que a imagem é dividida para o cálculo do COHAWES não podem ser tão pequenos que a quantidade de pixels não seja suficiente para que o valor da correção tenha significância.

¹<http://www.briancbecker.com/blog/research/pubfig83-lfw-dataset/>

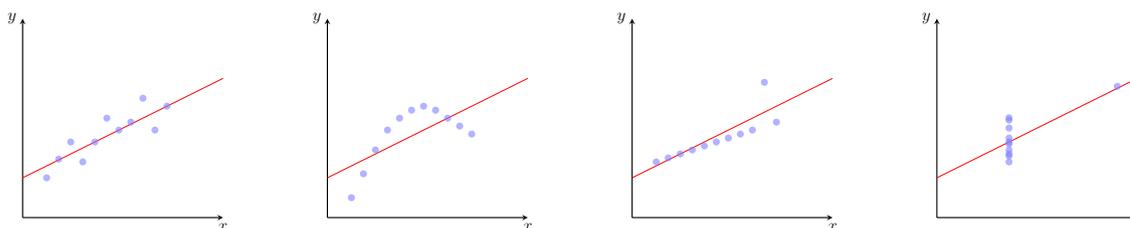


Figura 5.1: Diagramas de dispersão que possuem dispersões diferentes mas o valor da correlação é o mesmo.

5.4 DISPERSÃO DE DADOS DIFERENTES COM MESMA CORRELAÇÃO

Os diagramas de dispersão são gráficos que utilizam coordenadas cartesianas para exibir valores de um conjunto de dados, mas pares de variáveis com valores diferentes e dispersões diferentes podem gerar o mesmo coeficiente de correlação. Na Figura 5.1 são mostrados 4 exemplos de gráficos que possuem o mesmo valor de coeficiente de correlação. Com isso, existe a possibilidade de imagens visualmente diferentes gerarem os mesmos coeficientes de correlação durante o cálculo do descritor COHAWES, porém em sistemas de reconhecimento facial as imagens utilizadas são de domínio específico, fazendo com que apenas imagens de face sejam processadas pelo descritor, e nesse caso, quando os valores dos coeficientes foram iguais certamente são duas faces que estão sendo comparadas prove provavelmente as faces dessas imagens são parecidas.

Como trabalhos futuros, o descritor COHAWES poderá ser aplicado sistemas de reconhecimento de objetos onde os data sets possuem imagens de objetos variados e o descritor estará sujeito à valores de coeficientes próximos mas sem a garantia de que os objetos comparados são do mesmo tipo. Vale ressaltar que o COHAWES teve uma boa vantagem, em relação aos outros descritores, no reconhecimento de faces no data set PF83LFW que é data set com menos controle sobre as imagens dentre os data sets testados por este trabalho.

5.5 BENCHMARK PARA RECONHECIMENTO FACIAL

Outra contribuição deste trabalho é a implementação de um *framework* que auxiliou na configuração e execução dos cenários testados. Esse *framework* possibilita a parametrização de vários pontos da extração do COHAWES como: características do vetor de representação da imagem, tipo de relacionamento entre as características que será calculado (correlação ou covariância) e quantidade e tamanho das divisões da imagem que irão definir os *patches*. Além dos parâmetros do COHAWES esse *framework* permite a parametrização das etapas do reconhecimento facial permutando descritores, métricas de distância/classificadores e data sets. Apesar do uso da implementação disponibilizada por (ORTIZ; BECKER, 2014), para avaliar o descritor proposto nos vários classificadores, durante a integração entre os códigos foi necessária intervenção no código-fonte de Ortiz e Becker (2014) para executar os cenários. Como proposta de trabalhos futuros o

framework desenvolvido pode evoluir para uma ferramenta de *benchmark* onde o usuário não tenha necessidade de alterar o código da ferramenta seu e sim apenas inserir o código, ou configurar o data set, que deseja avaliar.

REFERÊNCIAS BIBLIOGRÁFICAS

- AHONEN, T.; HADID, A.; PIETIKAINEN, M. Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 28, n. 12, p. 2037–2041, 2006.
- AHONEN, T. et al. Face Recognition with Local Binary Patterns. In: *European Conference on Computer Vision*. [S.l.]: Springer, 2004. p. 469–481.
- ANAND, B.; SHAH, P. K. Face Recognition using SURF Features and SVM Classifier. *International Journal of Electronics Engineering Research*, v. 8, n. 1, p. 975–6450, 2016.
- AYINDE, O.; YANG, Y.-h. Face recognition approach based on rank correlation of Gabor-filtered images. *Pattern Recognition*, v. 35, n. 6, p. 1275–1289, jun 2002.
- BAY, H. et al. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, v. 110, n. 3, p. 346–359, jun 2008.
- BAY, H.; TUYTELAARS, T.; GOOL, L. V. SURF: Speeded Up Robust Features. In: LEONARDIS, A.; BISCHOF, H.; PINZ, A. (Ed.). *Computer Vision – ECCV 2006*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, (Lecture Notes in Computer Science, v. 3951). p. 404–417.
- BECKER, B. C.; ORTIZ, E. G. Evaluating open-universe face identification on the web. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, p. 904–911, 2013.
- BELHUMEUR, P. N.; HESPANHA, J. P.; KRIEGMAN, D. J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, IEEE, v. 19, n. 7, p. 711–720, 1997.
- BLEDSOE, W. W. The model method in facial recognition. *Panoramic Research Inc., Palo Alto, CA, Rep. PR1*, v. 15, p. 47, 1966.
- BOWYER, K. W.; CHANG, K.; FLYNN, P. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. *Computer Vision and Image Understanding*, v. 101, n. 1, p. 1–15, 2006.
- BROWN, M.; LOWE, D. Invariant Features from Interest Point Groups. *British Machine Vision Conference, BMVC 2002*, p. 656–665, 2002.
- CAMBRIDGE, A. The orl database of faces. *Cambridge, UK [Online]*, 2002.

CAMENT, L. A. et al. Face recognition under pose variation with local Gabor features enhanced by Active Shape and Statistical Models. *Pattern Recognition*, Elsevier, v. 48, n. 11, p. 3371–3384, 2015.

CORTES, C.; VAPNIK, V. Support-Vector Networks. *Machine Learning*, v. 20, n. 3, p. 273–297, 1995.

COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, v. 13, n. 1, p. 21–27, jan 1967.

CRISTIANINI, N.; SHAWE-TAYLOR, J. *An introduction to support vector machines and other kernel-based learning methods*. [S.l.]: Cambridge university press, 2000.

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*. [S.l.]: IEEE, 2005. I, p. 886–893.

DATTA, A. K.; DATTA, M.; BANERJEE, P. K. *Face Detection and Recognition :Theory and Practice*. [S.l.]: CRC Press, 2016. 342 p.

DÉNIZ, O. et al. Face recognition using Histograms of Oriented Gradients. *Pattern Recognition Letters*, v. 32, n. 12, p. 1598–1603, sep 2011.

DING, C. et al. Multi-Directional Multi-Level Dual-Cross Patterns for Robust Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 38, n. 3, p. 518–531, mar 2016.

DU, G.; SU, F.; CAI, A. Face recognition using SURF features. *Proceedings of SPIE*, v. 7496, n. 1, p. 749628–749628–7, 2009.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 2001. 680 p.

FIGUEIREDO, M. A. T.; NOWAK, R. D.; WRIGHT, S. J. Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems. *IEEE Journal of Selected Topics in Signal Processing*, v. 1, n. 4, p. 586–597, dec 2007.

GAN, J.; XIAO, J. A Weighted SRC Algorithm for Face Recognition with Disguise. v. 2, n. February, p. 513–520, 2012.

GEORGHIADES, A.; BELHUMEUR, P.; KRIEGMAN, D. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 23, n. 6, p. 643–660, jun 2001.

GOLDSTEIN, A. J.; HARMON, L. D.; LESK, A. B. Identification of human faces. *Proceedings of the IEEE*, IEEE, v. 59, n. 5, p. 748–760, 1971.

HE, X. et al. Face recognition using Laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 27, n. 3, p. 328–340, 2005.

- HEISELE, B.; HO, P.; POGGIO, T. Face recognition with support vector machines: global versus component-based approach. In: *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. [S.l.]: IEEE Comput. Soc, 2001. v. 2, p. 688–694.
- HUANG, G. B. et al. *Labeled faces in the wild: A database for studying face recognition in unconstrained environments*. [S.l.], 2007.
- KANADE, T. *Picture Processing System by Computer Complex and Recognition of Human Faces*. 143 p. Tese (Doutorado) — Kyoto University, 1973.
- KIRBY, M.; SIROVICH, L. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern analysis and Machine intelligence*, IEEE, v. 12, n. 1, p. 103–108, 1990.
- KUMAR, N. et al. Attribute and simile classifiers for face verification. *Proceedings of the IEEE International Conference on Computer Vision*, p. 365–372, 2009.
- LADES, M. et al. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, v. 42, n. 3, p. 300–311, 1993.
- LARSON, R.; FARBER, B. *Estatística Aplicada*. 4. ed. [S.l.]: Pearson Education, 2010. ISBN 978-85-7605-372-9.
- LI, B.; HUO, G. Face recognition using locality sensitive histograms of oriented gradients. *Optik - International Journal for Light and Electron Optics*, v. 127, n. 6, p. 3489–3494, 2016.
- LI, G. et al. An efficient face normalization algorithm based on eyes detection. In: *IEEE Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*. [S.l.], 2006. p. 3843–3848.
- LI, S. Z.; JAIN, A. K. (Ed.). *Face recognition applications*. 2nd. ed. [S.l.]: Springer, 2011.
- LI, Y.; QI, C. Face recognition using Hog feature and group sparse coding. In: *2013 IEEE International Conference on Image Processing, ICIP 2013 - Proceedings*. [S.l.: s.n.], 2013. p. 3350–3353.
- LIU, C.; WECHSLER, H. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, v. 11, n. 4, p. 467–476, 2002.
- MALIOUTOV, D.; CETIN, M.; WILLSKY, A. Homotopy continuation for sparse signal representation. In: *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005*. [S.l.]: IEEE, 2007. v. 5, n. 2, p. 733–736.
- MORETTIN, P. A.; BUSSAB, W. d. O. *Estatística Básica*. 6. ed. [S.l.]: Saraiva, 2010. ISBN 9788502081772.

NAGARAJU, C.; SRINU, B.; RAO, E. Recognition Under Difficult Lighting Conditions Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions. *International Journal of Innovative Technology and . . .*, p. 168–182, 2011.

NASEEM, I.; TOGNERI, R.; BENNAMOUN, M. Linear regression for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 32, n. 11, p. 2106–2112, 2010.

NORDSTRØM, M. M. et al. *The IMM Face Database - An Annotated Dataset of 240 Face Images*. Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 2004.

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

OJANSIVU, V.; HEIKKILÄ, J. Blur Insensitive Texture Classification Using Local Phase Quantization. In: . [S.l.]: Springer, Berlin, Heidelberg, 2008. p. 236–243.

ORTIZ, E. G.; BECKER, B. C. Face recognition for web-scale datasets. *Computer Vision and Image Understanding*, v. 118, p. 153–170, 2014.

OSBORNE, M.; PRESNELL, B.; TURLACH, B. A new approach to variable selection in least squares problems. *IMA Journal of Numerical Analysis*, v. 20, n. 3, p. 389–403, 2000.

OSUNA, E.; FREUND, R.; GIROSIT, F. Training support vector machines: an application to face detection. In: IEEE. *Computer vision and pattern recognition, 1997. Proceedings., 1997 IEEE computer society conference on*. [S.l.], 1997. p. 130–136.

PHILLIPS, P. et al. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 10, p. 1090–1104, 2000.

PHILLIPS, P. J. et al. The {FERET} evaluation methodology for face recognition algorithms. *IEEE Proceedings of Computer Vision and Pattern Recognition*, v. 22, n. 10, p. 137–143, 1997.

PINTO, N. et al. Scaling-up Biologically-Inspired Computer Vision: A Case-Study on Facebook. In: *IEEE Computer Vision and Pattern Recognition, Workshop on Biologically Consistent Vision*. [S.l.: s.n.], 2011.

RASCHKA, S. *About Feature Scaling and Normalization*. 2014. Disponível em: http://sebastianraschka.com/Articles/2014_about_feature_scaling.html.

SAMARIA, F. F.; HARTER, A. A. Parameterisation of a stochastic model for human face identification. In: *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*. [S.l.: s.n.], 1994. p. 2–6.

- SHI, Q. et al. Is face recognition really a Compressive Sensing problem? In: *CVPR 2011*. [S.l.]: IEEE, 2011. p. 553–560.
- SIROVICH, L.; KIRBY, M. Low-dimensional procedure for the characterization of human faces. *Josa a*, Optical Society of America, v. 4, n. 3, p. 519–524, 1987.
- STEGMANN, M. B.; ERSBØLL, B. K.; LARSEN, R. FAME – a flexible appearance modelling environment. *IEEE Trans. on Medical Imaging*, IEEE, v. 22, n. 10, p. 1319–1331, 2003.
- TAN, X. et al. Enhanced Local Texture Feature Sets for Face Recognition Under Different Lighting Conditions. v. 4778, p. 168–182, 2007.
- TAN, X.; TRIGGS, W. Fusing Gabor and LBP Feature Sets for Kernel-Based Face Recognition. *Analysis and Modeling of Faces and Gestures*, v. 4778, p. 235–249, 2007.
- TUZEL, O.; PORIKLI, F.; MEER, P. Region covariance: A fast descriptor for detection and classification. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, v. 3952 LNCS, p. 589–600, 2006.
- VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: IEEE. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. [S.l.], 2001. v. 1, p. I–I.
- VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. [S.l.: s.n.], 2004. v. 1, p. I–511–I–518.
- WANG, J. et al. Locality-constrained Linear Coding for image classification. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Francisco: IEEE, 2010. p. 3360–3367.
- WEINBERGER, K.; BLITZER, J.; SAUL, L. Distance metric learning for large margin nearest neighbor classification. *Advances in neural information processing systems*, v. 18, p. 1473, 2006.
- WISKOTT, L. et al. Face recognition by elastic bunch graph matching. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, v. 1296, n. July 1997, p. 456–463, 1997.
- WRIGHT, J. et al. Robust Face Recognition via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 31, n. 2, p. 210–227, feb 2009.
- ZHANG, N.; YANG, J. K Nearest Neighbor Based Local Sparse Representation Classifier. In: *2010 Chinese Conference on Pattern Recognition (CCPR)*. [S.l.]: IEEE, 2010. p. 1–5.

ZHANG, W. et al. Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In: *Proceedings of the IEEE International Conference on Computer Vision*. [S.l.]: IEEE, 2005. I, p. 786–791.

ZHAO, W. et al. Face recognition: A literature survey. *ACM computing surveys (CSUR)*, ACM, v. 35, n. 4, p. 399–458, 2003.