UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

DERZU OMAIA

UM SISTEMA PARA DETECÇÃO E RECONHECIMENTO DE FACE EM VÍDEO UTILIZANDO A TRANSFORMADA COSSENO DISCRETA

João Pessoa

2009

DERZU OMAIA

UM SISTEMA PARA DETECÇÃO E RECONHECIMENTO DE FACE EM VÍDEO UTILIZANDO A TRANSFORMADA COSSENO DISCRETA

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Centro de Ciências Exatas e da Natureza da Universidade Federal da Paraíba, como parte dos requisitos para a obtenção do título de Mestre em Informática.
Área de concentração: Sistemas Digitais (Processamento Digital de Imagens).

ORIENTADOR:

Prof. Dr. Leonardo Vidal Batista

João Pessoa

2009

O54d Omaia, Derzu.
Um sistema para detecção e reconhecimento de face em vídeo utilizando a transformada cosseno discreta / Derzu Omaia. - - João Pessoa: UFPB, 2009.
93 f.: il.
Orientador: Leonardo Vidal Batista.
Dissertação (Mestrado) – UFPB/CCEN.

1.Informática. 2.Processamento digital de imagens. 3.Reconhecimento de face. 4.Transformada Cosseno Discreta.

UFPB/BC *CDU: 004(043)*

Ata da Sessão Pública de Defesa de Dissertação de Mestrado do aluno ***Derzu Omaia****, candidato ao Título de Mestre em Informática na Área de Sistemas de Computação, realizada em 27 de agosto de 2009.*

Aos vinte e sete dias do mês de agosto do ano dois mil e nove, às quatorze horas, na Sala de Reuniões do Centro de Ciências Exatas e da Natureza da Universidade Federal da Paraíba, reuniram-se os membros da Banca Examinadora constituída para examinar o candidato ao grau de Mestre em Informática, na área de "*Sistemas de Computação*", na linha de pesquisa "*Processamento de Sinais e Sistemas Gráficos*", o Sr. Derzu Omaia. A comissão examinadora foi composta pelos professores doutores: Leonardo Vidal Batista (DI - UFPB), Orientador e Presidente da Banca Examinadora, Ed Porto Bezerra (DI-UFPB), como examinador interno e Edson J. R. Justino (PPGIA/PUC-PR), como examinador externo. Dando início aos trabalhos, o Prof. Prof. Leonardo Vidal Batista, cumprimentou os presentes, comunicou aos mesmos a finalidade da reunião e passou a palavra ao candidato para que o mesmo fizesse, oralmente, a exposição do trabalho de dissertação intitulado "*UM SISTEMA PARA DETECÇÃO E RECONHECIMENTO DE FACE EM VÍDEO UTILIZANDO A TRANSFORMADA COSSENO DISCRETA*". Concluída a exposição, o candidato foi argüido pela Banca Examinadora que emitiu o seguinte parecer: "*aprovado*". Assim sendo, deve a Universidade Federal da Paraíba expedir o respectivo diploma de Mestre em Informática na forma da lei e, para constar, eu, professora Tatiana Aires Tavares, coordenadora do Programa de Pós-Graduação em Informática, servindo de secretária, lavrei a presente ata que vai assinada por mim mesmo e pelos membros da Banca Examinadora. João Pessoa, 27 de agosto de 2009.

Tatiana Aires Tavares

Tatiana Aires Tavares

Prof. Dr. Leonardo Vidal Batista
Orientador (DI-UFPB)

Prof. Dr. Ed Porto Bezerra
Examinador Interno (DI-UFPB)

Prof. Dr. Edson J. R. Justino
Examinador Externo (PPGIA/PUC-PR)

Dedico este trabalho aos meus pais, Fábio Fernandes de Oliveira e Marluce Nobre de Mariz Maia, em especial a minha mãe, que sempre me incentivou e aconselhou a seguir vida acadêmica.

Agradecimentos

Aos meus pais, que sempre me incentivaram e deram o suporte necessário para meu desenvolvimento.

Ao meu orientador, Leonardo Vidal Batista, por acreditar no meu trabalho, pelo seu incentivo, atenção, confiança, amizade e excelente orientação.

Ao LAVID, em especial ao professor Guido Lemos de Souza Filho, por me acolherem durante todo o período do mestrado.

A JanKees van der Poel, por suas idéias, revisões e orientação.

A Marluce Nobre de Mariz Maia, minha mãe, pelas suas excelentes revisões no português, pelas imagens, e apoio geral.

A Hilda Nobre de Mariz Maia, minha tia, pelas revisões no inglês, dos artigos e abstracts.

A Flavia Maia Guimarães, pela revisão estrutural e do português.

Às pessoas que participaram do banco de dados UFPB, as quais eu não posso citar os nomes para manter o anonimato das mesmas.

Aos amigos, que por diversas vezes ficaram em segundo plano enquanto me dedicava ao mestrado, mas que sempre estiveram presentes nos momentos em que precisava.

Ao Programa de Pós-Graduação em Informática da UFPB (PPGI), seus professores e funcionários, pela oportunidade de realização desse trabalho.

Resumo

A face humana possui um padrão bastante complexo e variável, o que torna as operações de detecção e reconhecimento de face um problema desafiador. O campo de aplicação dessas operações é bastante abrangente, envolvendo principalmente aplicações de segurança, como autorização de acesso físico e lógico, rastreamento de pessoas e autenticação em tempo real. Além de aplicações de segurança, a detecção e o reconhecimento de faces também pode ser associado a outras aplicações, como interação homem-máquina e realidade virtual.

Diversos trabalhos de detecção e reconhecimento de face vêm sendo propostos e desenvolvidos pela comunidade científica, buscando continuamente uma maior precisão e eficiência. Atualmente já estão disponíveis detectores e reconhecedores de face com precisão superior a 95%. Sistemas comerciais também já estão disponíveis no mercado.

Este trabalho apresenta um estudo sobre os diversos métodos de detecção e reconhecimento de face existentes. Também foi analisada a possibilidade de desenvolvimento de um novo método de detecção de face utilizando Predição por Casamento Parcial (*Prediction by Partial Match,* PPM), Entropia e Transformada Cosseno Discreta (*Discrete Cosine Transform,* DCT). Propõe-se ainda, um novo método de reconhecimento de face baseado na DCT. Por fim, apresenta-se a arquitetura de um sistema de detecção e reconhecimento de face em vídeo. Para validação desta arquitetura, o sistema proposto foi implementado utilizando um dos melhores detectores encontrados na literatura e o reconhecedor produzido neste trabalho.

Diversos experimentos foram realizados e tanto o detector de face utilizado, quanto o reconhecedor desenvolvido mostraram-se eficientes, atingindo taxas de acerto compatíveis com os métodos mais atuais.

Palavras-chave: Detecção de Face, Reconhecimento de Face, Reconhecimento de Padrões, Processamento Digital de Imagens, Transformada Cosseno Discreta.

Abstract

Human face has a very complex and variable pattern, which makes the face detection and recognition operations a challenging problem. The scope of these operations is quite comprehensive, involving mainly security applications, such as authorization for physical and logical access, people tracking, and real time authentication. In addition to security applications, face detection and recognition can also be associated with other applications, such as human-computer interaction and virtual reality.

Several studies of face detection and recognition have been proposed and developed by researchers, pursuing greater precision and efficiency. Currently there are face detectors and recognizers with accuracy exceeding 95%. Commercial systems are available as well.

This work presents a study on several face detection and recognition methods. Also was discussed the possibility of developing a new face detection method using Prediction by Partial Match (PPM), Entropy and Discrete Cosine Transform (DCT). It is further proposed a new face recognition method based on DCT. Finally, is proposed an architecture for a face detection and recognition system in video. To validate the architecture, the proposed system was implemented using one of the best detectors in the literature and the recognizer produced in this work.

Several experiments were performed, and both the face detector used as the recognizer developed were effective, achieving success rates compatible with most current methods.

Keywords: Face Detection, Face Recognition, Pattern Recognition, Digital Image Processing, Discrete Cosine Transform.

Lista de Abreviaturas e Siglas

AC	⇒	*Alternating Current*
BMP	⇒	*Bitmap*
CMU	⇒	*Carnegie Mellon University*
DC	⇒	*Direct Current*
DCT	⇒	*Discrete Cosine Transform*
DWT	⇒	*Discrete Wavelet Transform*
FERET	⇒	*Facial Recognition Technology*
GLR	⇒	*Generalized Likelihood Ratio*
HMM	⇒	*Hidden Markov Models*
ICA	⇒	*Independent Component Analysis*
JPEG	⇒	*Joint Photographic Experts Group*
KLT	⇒	*Karhunen-Loève Transform*
KNN	⇒	*K-Nearest Neighbor*
LDA	⇒	*Linear Discriminant Analysis*
MIT	⇒	*Massachusetts Institute of Technology*
NN	⇒	*Nearest Neighbor*
ORL	⇒	*Olivetti Research Lab*
PCA	⇒	*Principal Components Analysis*
PPM	⇒	*Prediction by Partial Match*
RGB	⇒	*Red Green Blue*
SDRF	⇒	*Sistema de Detecção e Reconhecimento de faces*
SOM	⇒	*Self-Organizing Map*
SVM	⇒	*Support Vector Machines*

Índice de Equações

Equação:

Índice de Figuras

Índice

Introdução

A visão computacional permite aos sistemas digitais extraírem informações de imagens. Diversas informações podem ser extraídas, as quais podem ser utilizadas para o reconhecimento de padrões complexos, como texturas, objetos, textos, padrões biométricos, entre outros [32]. Essa capacidade de reconhecer padrões proporciona aos sistemas digitais um sistema de visão artificial ainda não tão eficiente quanto o humano, mas que já possui algumas características superiores, como visão noturna e *zoom*.

Diversos sistemas fazem uso da visão computacional. Exemplos destes podem ser encontrados na robótica, onde a visão computacional permite que robôs decidam seus próprios movimentos, e no uso militar, permitindo que aviões de guerra e mísseis acertem seus alvos com maior precisão. Sistemas de reconhecimento de padrões biométricos são largamente utilizados na área de segurança. Esses sistemas utilizam características humanas singulares, como impressões digitais, íris, voz e face, permitindo a diferenciação entre seres humanos [72].

A área de análise de faces pode ser dividida em diversas subáreas, como reconhecimento de face, detecção/localização de face, reconhecimento de expressões faciais e análise de poses [72]. É importante diferenciar detecção e reconhecimento. O reconhecimento de face consiste em identificar um indivíduo por intermédio da análise de sua face, comparando-a com outras faces pré-rotuladas. A detecção ou localização de faces é a determinação da presença e posição espacial de cada face existente em uma imagem. A detecção de face freqüentemente é utilizada como uma etapa inicial para o reconhecimento.

A operação de reconhecimento de faces pode ser abordada de duas formas: identificação e autenticação [1][32]. Na identificação, a face de uma pessoa é comparada a uma galeria de faces tendo como objetivo determinar a pessoa proprietária da face investigada. Na autenticação, a face investigada e seu suposto proprietário são previamente informados e o objetivo é certificar a informação declarada, comparando a face investigada com uma galeria de faces para confirmar que a mesma corresponde ao proprietário declarado [45].

O reconhecimento de faces possui um campo de aplicação abrangente, permitindo uma maior confiabilidade associada a aplicações de segurança [13][53][72]. Estas aplicações variam, incluindo desde a autorização de acesso físico ou virtual até as mais sofisticadas aplicações de rastreamento de suspeitos. Uma significante aplicação de segurança corresponde à autenticação em tempo real de portadores de documentos pessoais, como passaporte, carteira de identidade, carteira de motorista, cartão de crédito, seguridade social e outros. Além de aplicações de segurança, o reconhecimento de faces também pode ser associado a outras aplicações, como interação homem-máquina, programas de treinamento e realidade virtual [72].

A detecção de faces, isoladamente, pode ser utilizada como ferramenta de sensoriamento de tráfego humano, ou de vigilância automática. No sensoriamento, ela informa a freqüência de ocorrência de faces em uma imagem estática ou em um vídeo obtendo informações como a quantidade de faces que passam por segundo em determinado local. Como sistema de vigilância automático, permite que os sistemas emitam alertas, caso pessoas se aproximem, ou que armazenem, em maior resolução, as regiões de face em um vídeo.

A combinação da detecção com as outras subáreas de análise de faces permite uma gama enorme de aplicações. Combinada com o reconhecimento de face, por exemplo, possibilita que a autenticação em um sistema seja realizada de forma mais robusta, sem a necessidade do usuário ficar parado em uma posição fixa na frente de uma câmera específica. Combinada com o reconhecimento de face, permite ainda o rastreamento de pessoas em fotos e vídeos, possibilitando a busca, localização e identificação de pessoas específicas em grandes bancos de dados de vídeo, trabalho que seria praticamente inviável de ser feito manualmente por um ser humano. O rastreamento de faces pode ser utilizado, por exemplo, na área de segurança, buscando suspeitos; e na mídia televisiva, buscando algum ator em todos os vídeos da emissora. Combinando-se a detecção com a análise de expressões faciais é possível, por exemplo, determinar em uma máquina fotográfica se as pessoas na foto estão com os olhos fechados ou abertos, ou se estão sorrindo ou não.

A grande dificuldade para desenvolver detectores e reconhecedores de faces robustos é o fato de que as faces humanas não seguem um padrão rígido, variando muito em relação à forma, cor e tamanho. Ainda, o fato de uma mesma pessoa poder usar diferentes acessórios (como óculos, brincos, *piercings* e maquiagem); alterar o tamanho e estilo do seu cabelo; usar barba ou bigode; apresentar diversas expressões faciais; envelhecer; engordar ou emagrecer;

fazer cirurgias plásticas. Estas e outras alterações, dificultam ainda mais a tarefa de reconhecer faces de forma automática.

Estas variações tornam a tarefa de reconhecer ou detectar faces desafiadora, pois até mesmo seres humanos cometem enganos em alguns momentos. Desta forma, os métodos de reconhecimento e detecção de faces normalmente tentam abstrair as características mutáveis, concentrando-se nas características intrínsecas às faces [1][72]. Estas áreas de pesquisa estão em constante renovação, e novas tecnologias estão sempre surgindo. Atualmente existem métodos de detecção de face em vídeos de baixa resolução conseguindo taxas de acerto acima de 90% em tempo real [33][41][47][48][60][67]. Métodos de reconhecimento atuais também atingem taxas de acertos superiores a 90% [26][46][53] em tempo real e alguns atingem até 100% de acerto em certos testes, porém possuem um elevado custo computacional [9][39].

Este trabalho apresenta um estudo sobre os diversos métodos de detecção e reconhecimento de faces existentes. Também foi analisada a possibilidade de desenvolvimento de um novo método de detecção de face utilizando Predição por Casamento Parcial (*Prediction by Partial Match,* PPM), Entropia e Transformada Cosseno Discreta (*Discrete Cosine Transform,* DCT), contudo essas técnicas não se mostraram eficientes para o desenvolvimento de um detector de face com taxas de acerto compatíveis com as atuais [33][41][47][48][60][67][70]. Propõe-se, ainda, um novo método de reconhecimento de face baseado na DCT. Por fim, apresenta-se a arquitetura de um sistema de detecção e reconhecimento de face em vídeo. Para validação da arquitetura, o sistema proposto foi implementado utilizando um dos melhores detectores encontrados na literatura e o reconhecedor produzido neste trabalho [50][67].

O método de reconhecimento de faces desenvolvido neste trabalho é holístico e baseia-se na utilização dos coeficientes da DCT como atributos. Os coeficientes mais relevantes para o reconhecimento são selecionados utilizando um Seletor de Baixas Freqüências [45]. Durante a classificação, esses coeficientes são extraídos e, então, calcula-se a distância entre os coeficientes da face em análise para os coeficientes da galeria de faces do banco de dados utilizado. Entre todas as distâncias, a menor provavelmente será entre faces pertencentes a uma mesma pessoa, então a face em análise é classificada como pertencente a esta pessoa. Esta abordagem de classificação é conhecida por Classificador do Vizinho mais Próximo (*Nearest Neighbor*, NN) [45]. O cálculo da distância é realizado utilizando a técnica de *Minkowski* de ordem um. Essa distância foi utilizada por ser simples de calcular, sem

raízes nem quadrados, e por apresentar bom resultado como medida de similaridade entre as faces.

Foram realizados testes de classificação utilizando validação cruzada [25] sob o banco de dados do *Olivetti Research Lab* (ORL). Tais testes atingiram taxas de reconhecimento de 99,75% sem nenhum pré-processamento. Os resultados foram obtidos a um baixo custo computacional, uma vez que existem algoritmos eficientes tanto para a computação da DCT, quanto para o cálculo da distância.

O reconhecedor proposto pode ser considerado como uma continuação dos trabalhos de Hafed e Levine [26] e de Matos *et al.* [46], sendo que os resultados foram aprimorados e foram realizadas alterações nos procedimentos adotados para o reconhecimento. Esses trabalhos demonstram que o uso de coeficientes da DCT no reconhecimento de faces produz resultados com elevada taxa de acertos em menor tempo de processamento que outros métodos, sendo também relativamente independente de fatores como iluminação e posição. Isto se dá devido a propriedades da DCT de conseguir realizar uma redução da dimensionalidade dos dados, mantendo as características mais importantes [26][45].

1.1. Objetivos

O objetivo geral deste trabalho é o desenvolvimento e implementação de um sistema de detecção e reconhecimento de faces humanas em vídeo. Cada quadro do vídeo pode conter várias faces ou nenhuma; as faces podem ser de diferentes tamanhos, estar sob diferentes estados de iluminação, e possuir expressões faciais diversas.

Para atingir esse objetivo foram definidos os seguintes objetivos específicos:

1. Analisar os diversos métodos de detecção e reconhecimento de faces existentes;
2. Investigar a aplicabilidade do PPM, da DCT e da Entropia ao desenvolvimento de um novo método de detecção de face;
3. Investigar a aplicabilidade da DCT a um novo método de reconhecimento de face;
4. Analisar o uso de diversos filtros de pré-processamento de imagens que contribuam para o desenvolvimento desses métodos;
5. Desenvolver uma arquitetura de um sistema detector e reconhecedor de face;

6. Validar a arquitetura proposta;
7. Gerar um novo banco de faces em vídeo;
8. Realizar testes com bancos de faces referenciados na literatura; e
9. Comparar os resultados com os de outros métodos já publicados.

1.2. Estrutura da Dissertação

Esta dissertação está organizada em seis capítulos. Os tópicos a serem abordados em cada um dos capítulos estão descritos a seguir. Na introdução é realizada uma breve descrição do problema tratado, a motivação para o desenvolvimento, os principais objetivos, e a abordagem empregada na resolução do problema. O segundo capítulo apresenta a fundamentação teórica necessária ao desenvolvimento do trabalho. O terceiro capítulo descreve uma visão geral sobre reconhecimento e detecção de faces, e os trabalhos relacionados. O quarto capítulo apresenta os materiais e métodos propostos. O quinto capítulo descreve os resultados e comparações com outras técnicas. O último capítulo apresenta as conclusões e propõe trabalhos futuros. Por fim, é apresentado na forma de apêndice um artigo publicado durante o desenvolvimento deste trabalho.

Capítulo 2

Fundamentação teórica

Ao longo dessa dissertação serão utilizados conceitos fundamentais de Teoria da Informação, de Processamento de Imagens e de Análise de Sinais. Alguns desses conceitos são explicados a seguir.

2.1. Entropia

Na Termodinâmica, a entropia está relacionada à aleatoriedade (grau de desordem) das moléculas em um sistema. Na Teoria da Informação, a entropia (ou informação média) mede a incerteza ou a surpresa relacionada a um evento.

A informação associada a um símbolo *i* é definida na Equação (1), sendo P_i a probabilidade do símbolo *i* ocorrer. A entropia de ordem 1 contida em uma mensagem formada por *N* símbolos é definida pela Equação (2). A entropia de ordem 1 é utilizada quando considera-se que o símbolo atual tem probabilidade independente dos símbolos que o precedem [61]. Outras ordens também podem ser calculadas analisando-se a probabilidade relativa a símbolos anteriores, entretanto neste trabalho só utilizada a entropia de ordem 1.

$$I_i = \log_2\left(P_i^{-1}\right) \quad \textbf{(1)}$$

$$H = \sum_{i=0}^{N-1} P_i \log_2\left(P_i^{-1}\right) \quad \textbf{(2)}$$

2.2. Predição por Casamento Parcial

A Predição por Casamento Parcial (*Prediction by Partial Matching*, PPM) é um método avançado de compressão de dados, baseado em um modelo estatístico contextual adaptativo. O compressor atribui uma probabilidade condicionada ao contexto a cada símbolo gerado. Este, então, é codificado de acordo com essa probabilidade, que se altera dinamicamente no decorrer da compressão [61].

O modelo estatístico adaptativo mais simples conta a quantidade de vezes que cada símbolo ocorreu no passado e atribui ao símbolo atual uma probabilidade baseada neste passado. Considere-se, por exemplo, que 1000 símbolos tenham sido gerados pelo codificador até o momento, e que 30 deles foram a letra “q”. Se o próximo símbolo for “q”, ele será codificado com uma probabilidade estimada de 30/1000, e seu contador de ocorrências é incrementado em 1. Na próxima vez que um “q” for encontrado, será codificado com probabilidade de 31/t, sendo “t” o número total de símbolos gerados pelo codificador até o momento, não incluindo o último “q” [61].

Um tipo de modelo mais avançado é o contextual. Em um modelo contextual, a probabilidade para o símbolo S depende da freqüência de ocorrência do símbolo em contextos compostos pelos símbolos já codificados. A letra “u”, por exemplo, ocorre com probabilidade típica de aproximadamente 2% em textos da língua inglesa. Todavia, se o símbolo anterior for um “q”, a probabilidade de o próximo símbolo ser um “u” é cerca de 99%, pois o digrama “qu” é muito mais comum na língua inglesa que qualquer outro digrama iniciado com “q” [61]. Modelos contextuais são capazes de capturar de modo muito mais preciso a estrutura de mensagens complexas do que modelos não contextuais.

O tamanho máximo do contexto do PPM é limitado por restrições de *hardware*, como memória e velocidade de processamento. O número de possibilidades de combinações dos símbolos (possíveis contextos) cresce exponencialmente em relação ao tamanho do contexto. Por exemplo, em um contexto de tamanho 5, e utilizando símbolos de 8 bits (ou seja, alfabeto com 256 símbolos), a quantidade de contextos possíveis é de 256^5 = 1 TB. Generalizando, a quantidade de contextos é A^K, sendo A o tamanho do alfabeto e K o tamanho do contexto. Entretanto, um modelo real do PPM não utiliza toda essa memória. Por motivos práticos, apenas os contextos que ocorrem são armazenados em memória; os demais recebem uma probabilidade padrão [61].

Mesmo utilizando algoritmos otimizados, o PPM ainda não consegue se igualar em termos de velocidade de compressão aos formatos ZIP ou RAR bastante utilizados em aplicativos comerciais populares, como o WinZIP™. Entretanto, o PPM gera um modelo estatístico de excelente desempenho para a maior parte das mensagens de interesse prático. Com o avanço da velocidade dos processadores seu uso será bem mais difundido. As versões mais recentes do WinRAR™, 7-ZIP™, WizZIP™ já permitem optar pelo PPM, de forma a atingir razões de compressão mais elevadas.

2.3. Transformada Cosseno Discreta

A teoria das transformadas representa um papel dos mais importantes na área de processamento de sinais e imagens. As transformadas geram um conjunto de coeficientes a partir dos quais é possível restaurar as amostras originais do sinal.

Em muitas situações é conveniente aplicar-se uma operação matemática genericamente denominada de transformada sobre um sinal a ser processado, convertendo-o para o outro domínio (comumente o da freqüência), efetuar o processamento do sinal neste domínio e, finalmente, converter o sinal processado de volta ao domínio original.

Uma importante característica das transformadas refere-se a sua capacidade de gerar coeficientes descorrelacionados, concentrando a maior parte da energia do sinal em um reduzido número de coeficientes. Isso permite a redução da dimensionalidade dos dados analisados, conservando-se as características mais representativas do sinal [7].

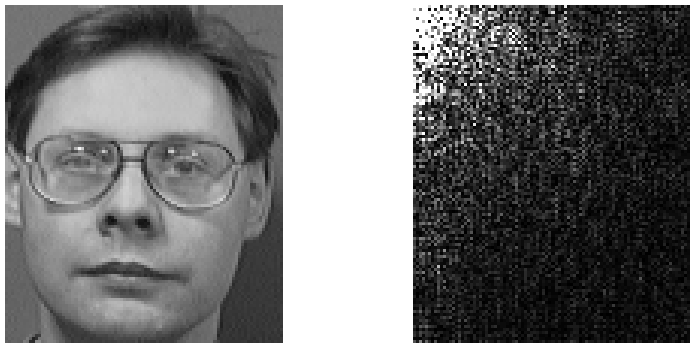
A Transformada Cosseno Discreta (DCT) é uma função linear e invertível, que expressa os sinais como uma soma de funções cosseno [55]. O sinal original é convertido para o domínio da freqüência e é possível converter o sinal de volta para o domínio original aplicando-se a DCT inversa.

Após a transformação para o domínio da freqüência, obtêm-se os coeficientes da DCT, que refletem a importância das freqüências presentes no sinal original. Os primeiros coeficientes referem-se às freqüências mais baixas do sinal, que representam as características gerais, normalmente as mais representativas do sinal original. Os últimos coeficientes referem-se às freqüências mais altas do sinal, que geralmente representam os detalhes, e as bordas ou o ruído presente no sinal [24]. Dessa forma, no caso específico de se

reduzir a dimensionalidade após a aplicação da DCT, os coeficientes de mais baixa freqüência são normalmente mais apropriados para representar os padrões de interesse.

Para o processamento de imagens, é interessante utilizar a DCT bidimensional (DCT-2D), visto que imagens são elementos bidimensionais. O padrão JPEG, por exemplo, estabelece o uso da DCT-2D na etapa de descorrelação [55]. No presente trabalho, utiliza-se a DCT-2D e, quando apenas o termo DCT é utilizado, subentende-se estar utilizando a DCT-2D.

A Figura 1 demonstra a aplicação da DCT sobre uma imagem de face do banco de dados ORL. A Figura 1.a mostra a imagem original, e a Figura 1.b apresenta o resultado da aplicação da DCT sobre a imagem original. Nela é possível perceber que a maior parte da energia está concentrada no início do sinal, na região superior esquerda da imagem. Essa região representa os coeficientes da DCT com as menores freqüências.



(a) Imagem original. (b) Imagem da DCT normalizada pessoa 4, em sua pose 1.

Figura 1: Imagem do banco de dados ORL, pessoa 4 em sua pose 1, e a DCT correspondente.

Para obtenção da Figura 1.b foi realizada uma normalização da matriz da DCT, visto que os coeficientes da DCT podem possuir valores inferiores ou superiores ao intervalo de níveis de cinza das imagens utilizadas, que variam de 0 a 255.

Nessa normalização, todos os coeficientes foram modularizados, ficando com valores positivos; em seguida identificou-se o maior valor de amplitude entre todos os coeficientes, desconsiderando o do nível DC (primeiro coeficiente, elemento [1,1] da matriz), porque ele geralmente possui valor substancialmente superior aos demais. Dividiram-se então, todos os coeficientes pelo valor máximo encontrado, e atribuiu-se 1 ao coeficiente DC. Desta forma todos os coeficientes ficam com valores entre 0 e 1, onde 0 corresponde a preto e 1 corresponde ao branco. Por fim, foi realizada uma equalização e uma expansão do histograma, para visualização mais apropriada dos coeficientes de alta freqüência.

A Figura 2 ilustra a reconstrução de uma imagem de face após a aplicação da DCT e da DCT inversa. A Figura 2.a corresponde à face original, de dimensão 92x112 pixels, ou seja, uma matriz de 10.304 valores. As duas faces seguintes representam a reconstrução da imagem original, utilizando, respectivamente, 2.576 e 625 coeficientes da DCT. Para se obter a imagem reconstruída, foi adotado o seguinte procedimento: aplicação da DCT sobre a face original, atribuição do valor zero aos coeficientes DCT a serem descartados e, por último, a aplicação da DCT inversa sobre a nova matriz de coeficientes. A Figura 2.b ilustra a reconstrução da face original considerando apenas os coeficientes de mais baixa freqüência da DCT. Foram selecionados os coeficientes do primeiro quadrante, ou seja, apenas 25% dos coeficientes DCT foram preservados: os coeficientes definidos por um retângulo de vértices [1,1], [1,46], [56,1] e [56,46] foram mantidos, e aos restantes foi atribuído o valor zero. Já a Figura 2.c ilustra a reconstrução da imagem original preservando apenas 6,07% dos coeficientes da DCT, ou seja, os coeficientes delimitados por um retângulo de vértices [1,1], [1,25], [25,1] e [25,25] foram preservados e aos restantes foram atribuídos o valor zero.

(a) Imagem original. (b) Imagem reconstruída preservando 25% dos coeficientes. (c) Imagem reconstruída preservando 6,07% dos coeficientes.

Figura 2: Imagem do banco de dados ORL, pessoa 4, em sua pose 1, e a reconstrução da imagem a partir dos coeficientes de baixa freqüência da DCT.

Pelas imagens reconstruídas da Figura 2, pode-se observar que a redução de dimensionalidade baseada em DCT produz bons resultados. Isto porque as imagens reconstruídas, considerando apenas os coeficientes DCT de baixa freqüência, apresentam redução de detalhes, mas preservam as informações importantes para caracterizar as imagens. Tais resultados sugerem ser viável um método de detecção ou reconhecimento de face que faça uso da redução da dimensionalidade baseada em DCT. Essa abordagem de redução de dimensionalidade já foi utilizada com sucesso por Matos *et al*. [46] e por Hafed e Levine [26] com o objetivo de fazer reconhecimento de faces.

Essa reconstrução da imagem, preservando apenas os coeficientes de mais baixa freqüência, funciona como um filtro passa-baixas ideal, ou seja, apenas as freqüências abaixo de um certo limiar são permitidas, e não há zona de transição das freqüências baixas para as altas. Por causa disto, ocorre o chamado fenômeno de Gibbs [23]. Este fenômeno gera oscilações nas imagens, como pode ser visto na Figura 2.b e na Figura 2c. Pretende-se, como trabalho futuro, investigar a aplicação de um filtro passa-baixas não ideal para minimizar o fenômeno de Gibbs.

A DCT-2D utilizada neste trabalho é a DCT-II, cuja definição é apresentada nas Equações (3) e (4) abaixo. Neste contexto, a imagem original corresponde à matriz de escala de cinza $x[m,n]$, de dimensões a x b. A aplicação da DCT-II produz então a matriz $X[k,l]$, também de dimensão a x b. As variáveis m e n são as coordenadas no domínio espacial e k e l são as coordenadas no domínio da freqüência [55].

$$X[k,l] = \frac{2}{N} c_k c_l \sum_{m=0}^{a-1} \sum_{n=0}^{b-1} x[m,n] \cos\left[\frac{(2m+1)k\pi}{2N}\right] \cos\left[\frac{(2n+1)l\pi}{2N}\right] \quad \textbf{(3)}$$

$$c_k, c_l = \begin{cases} \left(\frac{1}{2}\right)^{1/2} \text{ para } k = 0, l = 0 \\ 1 \text{ para } k = 1,2,...a-1 \text{ e } l = 1,2,...b-1 \end{cases} \quad \textbf{(4)}$$

O primeiro coeficiente, $X[1,1]$, é referenciado como sendo o coeficiente DC (*Direct Current*) e depende apenas do brilho médio da imagem. Os demais coeficientes de $X[k,l]$ indicam a amplitude correspondente do componente de freqüência de $x[m,n]$ e são referenciados como sendo os coeficientes AC (*Alternating Current*) [55].

2.4. Reconhecimento de padrões

Um padrão é algo que segue alguma regra, ou conjunto de regras, de forma que seja possível distingui-lo de outros padrões. Por exemplo, uma parede de tijolos, a areia do mar, um plantio de grama, uma impressão digital humana, um texto cursivo, uma face humana, entre diversos outros exemplos, apresentam um padrão típico. Reconhecimento de padrões é a capacidade de reconhecer e diferenciar os diversos padrões existentes [31].

O ser humano é um excelente reconhecedor de padrões, conseguindo detectá-los com alta qualidade e rapidez [45]. O cérebro humano é bem adaptado a essa função, pois durante todo o processo evolutivo humano, a habilidade de detectar padrões foi decisiva para a sobrevivência da espécie. Essa habilidade permitiu diferenciar os diversos padrões vistos, como predadores, presas e toda a natureza ao seu redor. Já a habilidade humana para cálculos é bastante limitada quando comparada a uma simples calculadora. Os computadores, ao contrário dos seres humanos, possuem uma imensa habilidade para fazer operações matemáticas. Por outro lado, os sistemas computacionais, a princípio, não são bons reconhecedores de padrões.

Existem diversos métodos de reconhecimento de padrões, dentre os quais quatro se destacam: casamento de padrões, casamento sintático ou estrutural, redes neurais e classificação estatística [31]. O estudo sobre classificação estatística será mais aprofundado, visto que esta é a abordagem utilizada neste trabalho.

2.4.1. Casamento de Padrões

Casamento é uma operação genérica, no reconhecimento de padrões, que é utilizada para determinar a similaridade entre dois objetos do mesmo tipo. No casamento de padrões, o objeto em análise é comparado a padrões previamente armazenados, através de alguma função de similaridade, como distância ou correlação. A função de similaridade deve ser otimizada de acordo com o banco de dados de treinamento. O resultado dessa função é comparado a algum limiar. Caso o resultado seja inferior ao limiar, o objeto pertence à classe; se superior, ele não pertence [31].

O casamento de padrões é eficiente para alguns tipos de aplicações, onde os objetos a serem classificados têm poucas distorções. Porém, em um domínio de aplicações onde os objetos variem em relação a algum processamento na imagem, mudança do ponto de vista ou possuam grandes variações intra-classe, o classificador pode não atingir bons resultados [31].

2.4.2. Casamento Sintático

No casamento sintático cada padrão é formado por uma combinação de diversas unidades elementares, chamadas de primitivas. É realizada uma analogia entre a estrutura do padrão e a sintaxe de uma linguagem, onde os padrões são vistos como sentenças da

linguagem, geradas de acordo com uma gramática, e as primitivas são o alfabeto da linguagem. Assim, uma grande variedade de padrões complexos pode ser descrita por um pequeno número de primitivas e regras gramaticais. A gramática de cada padrão deve ser construída a partir da formação das amostras desse padrão. A decisão é tomada verificando a probabilidade do objeto em análise pertencer a alguma gramática de padrão [31].

Essa abordagem é utilizada em situações onde os padrões têm uma estrutura bem definida a qual pode ser extraída em termos de um conjunto de regras. Por exemplo, eletrocardiogramas, imagens com texturas bem definidas, e análise de contorno de formas.

2.4.3. Redes Neurais

O cérebro humano é capaz de processar uma quantidade muito grande de informações rapidamente. Pesquisas em inteligência artificial procuram organizar elementos processadores de forma similar à organização dos neurônios do cérebro humano, buscando obter uma capacidade de processamento similar. Analogamente ao cérebro humano, uma rede neural artificial é composta por elementos processadores chamados neurônios, densamente interconectados por múltiplas conexões ponderadas, que são capazes de adquirir conhecimento com o passar do tempo. Os neurônios artificiais adquirem conhecimento através de seu relacionamento com os demais neurônios, baseando-se na repetição de um conjunto de soluções onde a saída de um neurônio da rede compõe a entrada de outro [45].

No reconhecimento de faces baseado em redes neurais, a rede é treinada para reconhecer certo padrão através de uma função não linear. A tomada de decisão é realizada comparando os resultados da função do objeto em análise com os resultados típicos de cada padrão [31].

2.4.4. Classificação Estatística

Na classificação estatística cada padrão é representado por *d* atributos formando um ponto em um espaço *d*-dimensional. Os atributos devem ser selecionados de forma que os pontos distribuam-se em regiões separadas (idealmente disjuntas) do espaço de atributos, de acordo com as classes de padrões de interesse. Essas regiões são delimitadas analisando-se a distribuição probabilística dos padrões neste espaço. A tomada de decisão é realizada

verificando se o ponto, que representa o objeto em análise, está contido em alguma dessas regiões [31].

Este método de classificação é dividido em duas etapas principais: treinamento (ou aprendizagem) e classificação. Essas etapas possuem sub-etapas de pré-processamento e seleção de atributos [45]. A Figura 3 detalha este método.

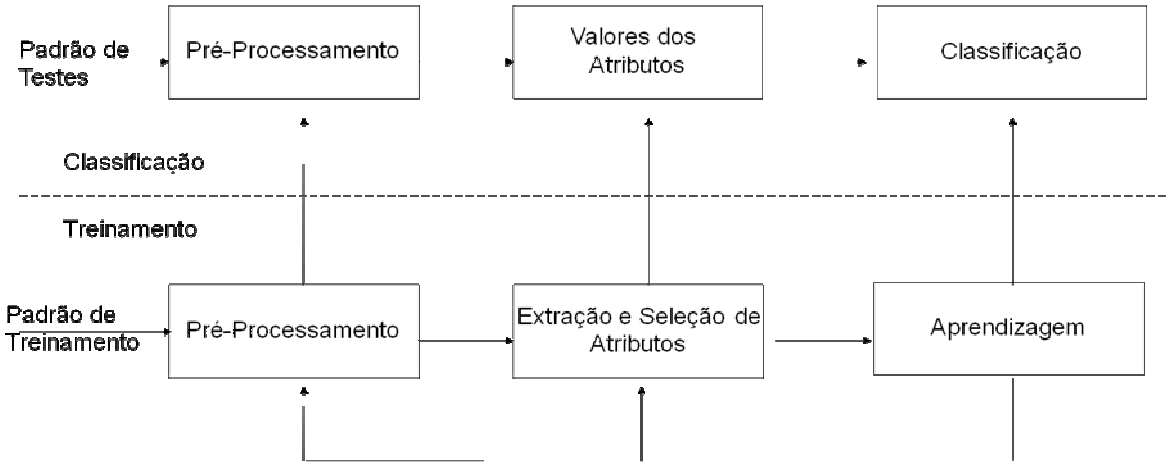


Figura 3: Sistema de reconhecimento estatístico, adaptado de Jain [31].

O treinamento é uma etapa de aprendizado, onde os atributos são selecionados e combinados de forma a representarem os padrões relevantes adequadamente. Na classificação são tomadas as decisões sobre a que classe de padrões pertence um dado padrão desconhecido [16]. O módulo de pré-processamento visa uma normalização entre as imagens, permitindo uma representação compacta e robusta do padrão. Com este objetivo são realizados uma série de tratamentos: balanceamento e equalização de brilho e iluminação, remoção de ruído e eliminação de ambiente e paisagem [45].

Quando em modo de treinamento, as principais características (atributos) dos padrões de interesse são selecionadas. Esses atributos devem ser selecionados de forma que possuam uma pequena variação intra-classe (entre diferentes amostras da mesma classe de padrões), e uma grande variação inter-classe (entre amostras de classes de padrões diferentes). A etapa de treinamento pode ser cíclica, onde a cada interação realiza-se um aperfeiçoamento do seletor de atributos.

Quando em modo de classificação, os atributos são apenas lidos do objeto em análise e comparados aos atributos pré-extraídos dos padrões. De acordo com a distância entre o objeto (um ponto no espaço de atributos) e os objetos ou classes de treinamento, decide-se a qual padrão pertence o objeto.

O problema de representar dados em um espaço *d*-dimensional é que, para *d* elevado, a representação dos dados pode ser ineficiente, além de ser mais custosa computacionalmente [31].

O desempenho de um classificador depende da inter-relação entre a quantidade de atributos selecionados (dimensão do espaço) e a quantidade de amostras das classes. Intuitivamente, pode-se pensar que quanto mais atributos melhor será a representação das classes. Porém, esse comportamento é observado apenas enquanto a quantidade de atributos é substancialmente menor que a quantidade de amostras das classes utilizadas no treinamento. Quando esta proporção é invertida, passa-se a ter uma degradação no desempenho do classificador. Este comportamento é conhecido como o Problema da Dimensionalidade [31].

A Figura 4 ilustra uma curva típica associada ao Problema da Dimensionalidade. Na região entre 0 e m_1 ocorre o comportamento esperado, ou seja, com o aumento da quantidade de atributos ocorre também aumento na eficácia do classificador. Entre m_1 e m_2, ocorre uma certa estabilidade, pois o aumento na dimensionalidade praticamente não influencia na taxa de acerto. Esta estabilidade sugere que os atributos importantes já foram considerados. Após m_2 inicia-se uma redução da eficácia do classificador, o que sugere que os atributos extras não são relevantes, e podem ser considerados ruído. O ideal para um classificador é trabalhar sobre o ponto m_1, ou seja, a menor dimensionalidade que ofereça a maior taxa de acerto.

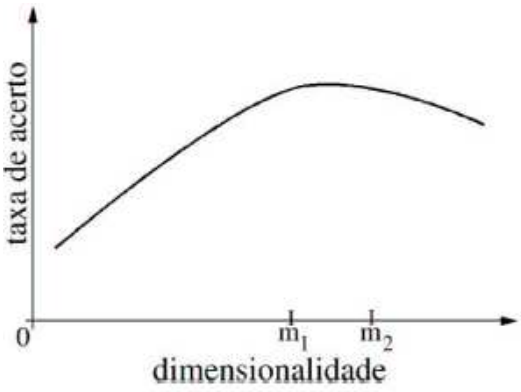


Figura 4: Problema da dimensionalidade, adaptado de Campos [12].

Capítulo 3

Detecção e Reconhecimento de Faces

Detecção e reconhecimento de faces são duas tarefas distintas, mas que comumente são utilizadas em conjunto. Essas áreas de pesquisa tornaram-se largamente estudadas nos últimos anos, e diversos métodos surgiram. Normalmente esses métodos adotam um conjunto de atividades comuns que os conduzem a um sistema robusto e com maior número de acertos [45].

Um processo típico de detecção e reconhecimento de faces normalmente estabelece a execução das seguintes atividades: normalização, extração de características, comparação com o banco de dados ou modelo, e decisão final [30]. A normalização compensa variações que possam existir em uma face, tratando em especial aspectos como iluminação, aproximação e posição, produzindo ao final uma imagem o mais próxima possível do padrão do banco de dados da comparação. A extração de características (seleção de atributos) gera o conjunto de atributos que serão utilizados no processo de comparação com o banco de dados. A comparação com o banco de dados verifica o grau de coincidência entre o conjunto de atributos selecionados da face desconhecida e os mesmos atributos das imagens armazenadas no banco de dados. Essa comparação também pode ser realizada diretamente com um modelo de faces do banco de dados. A decisão final conclui sobre o reconhecimento considerando as comparações efetivadas [45]. Tanto a comparação quanto a decisão final pertencem à etapa de classificação.

3.1. Seleção de Atributos

A etapa de seleção de atributos especifica a lista de características que representam uma determinada classe de padrão (no caso faces). Idealmente, selecionam-se atributos que

sejam similares entre objetos de uma mesma classe, substancialmente diferentes entre objetos de classes distintas. Em outras palavras, a seleção de atributos visa selecionar atributos que possuam uma pequena variação intra-classe e uma grande variação inter-classe [16]. Quando em modo de detecção, são selecionadas as características comuns a todas as faces e que permitam diferenciá-las dos demais padrões (não-faces). Neste caso, apenas dois padrões são comparados, o padrão de face e o padrão de não-face. Quando em modo de reconhecimento, são selecionadas as características particulares a cada pessoa, que permitam diferenciar sua face das faces da outras pessoas.

Fazendo-se uma abstração em alto nível, no modo de detecção essas características poderiam ser: dois olhos, um nariz e uma boca. Todavia, no modo de reconhecimento essas características não são suficientes para diferenciar pessoas, pois quase todas possuem essas características. Neste caso, características mais específicas teriam que ser utilizadas, como: cor dos olhos, formato do nariz, tamanho da boca, formato do rosto, testa, cor de pele, entre outras.

Uma imagem monocromática de face pode ser considerada como uma matriz 2D de h linhas por w colunas, e pode ser representada como um padrão no espaço de imagens de dimensionalidade $N = w * h$. Considerando que imagens digitais podem possuir alta resolução, o valor de N pode ser muito elevado. Porém, devido ao problema da dimensionalidade e do custo computacional, um valor elevado de N não é apropriado. Adicionalmente, muitos dos atributos podem ser irrelevantes ou até prejudiciais para a classificação. Por isso, comumente é realizada uma redução da dimensionalidade dos atributos, de forma a obter um grupo de atributos conciso e representativo das faces.

Para reduzir a dimensionalidade dos dados é necessário realizar uma seleção ou extração dos atributos mais relevantes. A seleção de atributos escolhe o conjunto mais representativo de atributos dentre os atributos originais. Já a extração, cria novos atributos a partir de transformações ou combinações dos atributos originais e seleciona o conjunto de atributos mais relevantes após essas transformações. Entretanto, comumente essas duas técnicas são tratadas como sinônimos na literatura [12].

Os métodos de seleção/extração de atributos transformam o espaço d-dimensional dos atributos em um espaço n-dimensional, onde $n \leq d$ [12]. Existem métodos lineares e não lineares capazes de realizar essa extração.

Os métodos lineares caracterizam-se por aplicarem uma mudança de base sobre o espaço original dos atributos, permitindo conseqüentemente a inversão da transformação realizada. As transformadas DCT, KLT e *Wavelet* transformadas lineares, e métodos lineares que utilizam essas transformadas têm-se mostrado eficientes no reconhecimento [6][9][26][46][50][53][65] e detecção de face [37][54]. Já os métodos não-lineares realizam uma transformação que não é inversível; uma vez realizada ela não pode ser desfeita. Estes últimos geralmente são baseados em redes neurais.

O método de seleção de atributos utilizado no reconhecedor de faces desenvolvido neste trabalho é o Seletor de Baixas Freqüências, e será explicado a seguir.

3.1.1. Seletor de Baixas Freqüências

O Seletor de Baixas Freqüências seleciona apenas os coeficientes correspondentes às freqüências mais baixas de um sinal [45]. O uso desse seletor em associação com a DCT, por exemplo, justifica-se porque esta transformada possui a propriedade de concentrar a maior parte da energia do sinal em um pequeno número de coeficientes de baixa freqüência [7].

A abordagem de seleção por Baixas Freqüências é simples, já que não analisa os valores dos coeficientes, nem realiza quaisquer cálculos nem comparações com eles. Apenas as posições dos coeficientes são consideradas, o que torna o processo de seleção simples e eficiente.

A Figura 5 ilustra a seleção de atributos através de regiões quadradas sobre uma imagem de DCT. As regiões quadradas delimitam os coeficientes que serão selecionados. Pode-se perceber que a maior concentração de energia ocorre nos coeficientes de mais baixa freqüência, onde estão os quadrados de seleção. Desta forma esta abordagem consegue capturar importantes coeficientes da imagem.

Figura 5: Regiões quadradas da seleção de baixas freqüências sobre uma DCT.

Outras formas geométricas também podem ser aplicadas sobre a DCT para a seleção dos atributos. Neste trabalho, formas geométricas retangulares, triangulares e elípticas foram avaliadas.

3.2. Abordagens de Classificação

Um classificador indica a que classe pertence determinada imagem de teste [12]. Quando no modo de detecção de face, o classificador indica se a imagem é uma face ou uma não-face. Quando no modo de reconhecimento, ele indica a qual pessoa pertence a face de teste.

Existem abordagens de classificação supervisionadas e não-supervisionadas. Na classificação com aprendizado supervisionado, amostras de todas as classes a serem classificadas são previamente definidas, e o classificador inicia o treinamento tendo um conhecimento prévio das classes que irá reconhecer. [69]. A maioria dos métodos de reconhecimento de faces utiliza a classificação supervisionada, onde as pessoas a serem reconhecidas têm suas imagens de faces separadas em classes [15].

Na classificação com aprendizado não-supervisionado, as classes a serem reconhecidas são inicialmente desconhecidas. Então, baseando-se nas similaridades entre os padrões, tenta-se reconhecer automaticamente, sem a intervenção humana, as possíveis classes existentes [69].

O Classificador de Bayes é um classificador ótimo do ponto de vista estatístico, porém ele possui exigências que nem sempre podem ser cumpridas em um sistema prático de reconhecimento e detecção de face. O classificador de Bayes exige que se saiba da

probabilidade a priori P_i e da probabilidade condicional $p(x|w_i)$ [12]. Embora existam métodos para estimação destas probabilidades, o custo computacional para a obtenção de uma representação precisa é alto. Em sistemas de detecção e reconhecimento de faces geralmente não se aplica a regra de decisão de Bayes, sendo utilizados, como alternativas, classificadores baseados em similaridade, como: K-Vizinhos mais Próximos, Vizinho mais Próximo e Distância Mínima ao Protótipo.

3.2.1. Classificador dos K-Vizinhos Mais Próximos

O Classificador dos K-Vizinhos Mais Próximos (*K-Nearest Neighbors*, KNN) calcula uma distância da imagem de padrão a ser classificada para todas as amostras de padrões do banco de treinamento. As K imagens com menores distâncias para a imagem de teste são selecionadas, e entre elas, a classe que ocorrer com maior freqüência será considerada como a que contém o padrão da imagem de teste [12].

Formalmente, sejam $\{y_{1j}, y_{2j}, ..., y_{mj}\}$ os m coeficientes DCT selecionados para representar a classe j e sejam $\{w_{1jk}, w_{2jk}, ..., w_{mjk}, \}$ as amplitudes dos coeficientes DCT de treinamento da classe j na amostra n, onde w_{ijk} corresponde ao coeficiente de mesma posição que y_{ij} [45].

Seja f uma imagem a ser classificada e sejam $\{v_{1f}, v_{2f}, ..., v_{mf}\}$ as amplitudes dos coeficientes DCT da imagem f, com v_{if} correspondendo ao coeficiente de mesma posição que y_{ij} [45].

O classificador KNN classifica a imagem f baseada nos seguintes passos:

1. Calcula-se a distância (DKNN) entre a imagem f e a classe de treinamento j na amostra n, com j=1, 2, ..., p e n=1, 2, ..., q, dada por:

$$DKNN_{fjk} = \sum_{i=1}^{m} \left| w_{ijk} - v_{if} \right| \qquad \textbf{(5)}$$

   Onde p é a quantidade de amostras para cada classe, e q é a quantidade de classes que o sistema classificará.

2. Identificam-se os K menores valores $DKNN_{fjk}$;

3. A classificação da imagem *f* corresponde à classe *j* mais freqüente entre os K-vizinhos identificados;
4. Em caso de empate, pode-se selecionar a classe com a menor das distâncias.

3.2.2. Classificador do Vizinho Mais Próximo

O Classificador do Vizinho Mais Próximo é um caso particular do Classificador dos k-Vizinhos mais próximos, quando k é igual a um [12]. Intuitivamente, pode-se pensar que o classificador dos K-Vizinhos mais próximos seria superior ao que analisa apenas o primeiro vizinho mais próximo, entretanto em diversos casos [26][45][50] este último classificador obteve melhores resultados.

3.2.3. Classificador de Distância Mínima ao Protótipo

O Classificador de Distância Mínima ao Protótipo consiste na definição de protótipos, no mínimo um para cada classe. Esses protótipos são vetores no espaço de atributos, que representam as classes. Uma forma comum de obtenção do protótipo de uma classe é através da média (baricentro) de suas amostras [12].

Após a geração dos protótipos, o classificador comporta-se como um classificador dos K-Vizinhos mais Próximos, onde, no lugar das amostras dos padrões, têm-se protótipos dos padrões [12].

3.3. Validação Cruzada

Para validar um classificador deve-se testar a sua eficácia. Uma técnica muito utilizada para essa validação, quando se tem um número reduzido de amostras para treinamento e testes, é a validação cruzada. Para tal, o banco de dados utilizado é dividido em dois grupos disjuntos, um para treinamento e outro para testes de classificação. Várias rodadas de classificação são realizadas com diferentes divisões dos grupos, a fim de se obter um resultado médio [25].

Um caso especial da Validação Cruzada é a técnica do deixe-um-de-fora (*leave-one-out*). Como o próprio nome sugere, essa técnica deixa uma amostra de cada classe fora do conjunto de treinamento. Esta será a amostra de teste ou de classificação, e o treinamento é

realizado com as demais. O processo é repetido N vezes, onde N é a quantidade de amostras no banco de dados, até que todas as amostras do banco tenham ficado de fora do treinamento. A média dos resultados de classificação destas repetições será o resultado final [25].

A Figura 6 ilustra a abordagem do deixe-um-de-fora, onde a cada rodada um elemento (em cinza) é classificado, e os demais elementos (em branco) são utilizados no treinamento. A cada rodada modifica-se o conjunto de treinamento e o elemento a ser classificado, até que todos os elementos tenham ficado uma vez de fora.

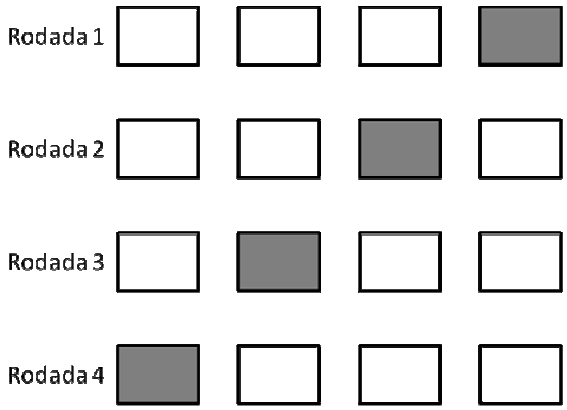


Figura 6: Deixe-um-de-fora, adaptado de Sirovich [62].

3.4. Métodos de Reconhecimento de Face

Os primeiros trabalhos de reconhecimento de faces, como os de Kelly [36] e o de Kanede [35], datam da década de 1970, e são baseados na relação entre a posição e tamanho dos atributos básicos em um rosto, como olhos, nariz, boca e orelhas [72]. Contudo, nesses trabalhos, essas regiões de atributos eram definidas manualmente. A partir de então, diversos outros métodos foram propostos, os quais podem ser divididos basicamente em três categorias: não holísticos, holísticos, e híbridos. Os métodos não holísticos se baseiam em características físicas individuais como olhos, nariz, boca e orelha. Já os métodos holísticos analisam a face como um todo, sem identificar características físicas individuas, e têm produzido resultados eficientes, visto que pequenas diferenças nas imagens comparadas não prejudicam o reconhecimento como um todo. A combinação dos métodos holísticos com os baseados em características gera os métodos híbridos.

Diversos métodos foram utilizados no desenvolvimento dos métodos propostos na literatura. Dentre eles se destacam os métodos baseados em transformadas (principalmente KLT, a partir da qual o método das *Eigenfaces* [65] é derivado, *Wavelets* e DCT); aqueles

baseados nas relações geométricas entre os atributos da face, os que utilizam informação 3D, aqueles que atuam sobre vídeo; os que utilizam redes neurais, entre outras.

As subseções seguintes contemplam algumas abordagens relevantes adotadas por métodos de reconhecimento de faces da atualidade. Métodos baseados em KLT/PCA e DCT são discutidos em seções específicas por apresentarem um maior grau de relacionamento com o método proposto neste trabalho. Outras abordagens com significantes contribuições para a área são referenciadas conjuntamente na seção 3.4.3.

3.4.1. Métodos Baseados na Transformada de *Karhunen-Lòeve*

A Transformada de Karhunen-Lòeve (*Karhunen-Lòeve Transform*, KLT) é uma transformada matemática. Transformadas são operações matemáticas invertíveis que convertem um sinal do domínio em que ele se encontra para outro domínio, por exemplo, do domínio do espaço para o domínio da freqüência. A utilidade da transformada para o reconhecimento/detecção de faces depende de sua capacidade para gerar coeficientes descorrelacionados e de concentrar a maior parte da energia do sinal em um pequeno número de coeficientes. A transformada pode ser vista como um pré-processamento que explora a correlação entre os pixels de uma imagem, permitindo a extração dos atributos mais relevantes a partir dos coeficientes de maior energia [7].

Turk e Pentland [65] propuseram um método de reconhecimento de faces não holístico batizado de *Eigenfaces*, motivado pelo trabalho de Sirovich e Kirby [62]. Neste método, as *Eigenfaces* são consideradas como um espaço N-dimensional de atributos de faces composto por autovetores (*eigenvectors*), os quais são vetores que contêm os atributos de face mais relevantes autovalores (*eigenvalues*) [65]. As *Eigenfaces* são geradas a partir da transformada KLT, a qual gera um novo espaço de atributos com a mesma dimensão do espaço original. Neste espaço, os atributos são mais facilmente representados. Essa transformação comumente é seguida por uma redução de dimensionalidade através da seleção das componentes principais da KLT. Os autovetores são obtidos a partir da matriz de covariância, *n*-dimensional, da distribuição probabilística dos *n*-atributos de uma face [25].

O trabalho de Turk e Pentland obteve taxas de acerto de 96%, sobre um banco de dados particular contendo 2500 faces [65]. Após esse trabalho surgiram diversos outros também baseados em PCA e *Eigenfaces*. Bartlett *et al.* [6] propõem um método baseado na

análise das componentes independentes (*Independent Component Analysis*, ICA), uma generalização da PCA, atingindo uma taxa de acerto de 87% sobre o banco de dados FERET. Ruiz-Del-Solar e Navarrete [58] realizaram testes sobre os bancos de dados FERET e Yale, atingindo taxas de 95,7% e 83,3% respectivamente.

Quando o sinal se comporta como um processo estocástico do tipo Markov-1, pode-se mostrar que a KLT é ótima no sentido que a mesma gera uma representação para os dados através da qual é possível reduzir-se a dimensionalidade dos mesmos com o menor erro médio quadrático. No entanto, a implementação da KLT é elaborada, exigindo a estimação da matriz de autocovariância do sinal a ser comprimido e sua diagonalização, bem como o cálculo da transformada propriamente dita, de forma que, na maioria dos casos, transformadas sub-ótimas são utilizadas. A DCT e a DWT são sub-ótimas e estão entre as transformadas mais empregadas em esquemas práticos de compressão de dados [7].

3.4.2. Métodos Baseados na Transformada Cosseno Discreta

Trabalhos de reconhecimento de face baseados na DCT se mostram mais rápidos do que os baseados em KLT/PCA, visto que a DCT possui cálculo mais simples e seus resultados são bem próximos aos da KLT [7][17].

Podilchuk e Xiaoyu [53] propuseram um método de reconhecimento de faces não holístico que define blocos posicionados sobre áreas expressivas da face humana, como olhos e boca. O método define os blocos representativos das principais características da face humana, aplica a DCT sobre tais blocos para cada uma das imagens de treinamento e finaliza com a classificação por distância mínima. Este método atingiu uma taxa de acerto de 94% sobre um banco de faces particular de 500 imagens.

O método de reconhecimento de faces de Hafed e Levine [26] é holístico e aplica a DCT tanto sobre as faces de treinamento quanto sobre a face de teste, seleciona os 49 coeficientes da DCT de mais baixa freqüência de cada face e aplica o método do vizinho mais próximo para classificar a face de teste em relação a todas as faces de treinamento, considerando apenas 49 coeficientes selecionados. A taxa de acerto obtida foi de aproximadamente 92,5% sobre o banco de dados ORL. Essa taxa foi atingida no caso especifico em que o treinamento foi realizado com as cinco primeiras faces do banco e os testes com as outras cinco faces restantes.

Matos *et al.* [46] propôs um método holístico que também aplica a DCT tanto sobre as faces de treinamento quanto sobre a face de teste, entretanto em sua abordagem foram realizados testes com diversos seletores de atributos e diversos classificadores. Os melhores resultados foram obtidos utilizando um seletor por baixas freqüências e um classificador do vizinho mais próximo utilizando apenas os 36 primeiros coeficientes. Neste caso, foi atingida uma taxa de reconhecimento de 99,25% sobre o banco de dados ORL.

O presente trabalho propõe um método de reconhecimento de faces baseado em DCT que possui resultados superiores aos acima mencionados, atingindo uma taxa de reconhecimento de 99,75% sobre o banco de dados ORL.

3.4.3. Métodos Adicionais

Outras transformadas também são utilizadas em trabalhos de reconhecimento de faces, como a DWT e LDA, bem como os métodos híbridos.

A Análise de Discriminantes Lineares de Fisher (*Linear Discriminant Analysis*, LDA), conhecida como *Fisherfaces* quando aplicada a reconhecimento de faces, é uma abordagem que extrai linearmente as características mais discriminantes das classes existentes a partir das informações associadas a cada padrão. A separação interclasses é enfatizada através da substituição da matriz de covariância adotada pelo PCA pela medida de separação de Fisher [18]. Belhumeur [8] propôs um reconhecedor de faces holístico baseado em LDA que atingiu uma taxa de acerto de 94% sobre o banco de dados Yale.

A transformada *wavelet* discreta (*Discrete Wavelet Transform*, DWT), é uma transformada matemática, assim como a DCT e KLT, que também gera coeficientes descorrelacionados. Chien e Wu [14] propõem um método holístico que utiliza a DWT para gerar vetores de atributos e sobre esses vetores é aplicada a LDA. Utilizando um classificador de vizinho mais próximo (*Nearest Neighbor*, NN) foi obtida uma taxa de reconhecimento de 94,5% sobre o banco de dados ORL.

Bicego et al. [9] propuseram um método não holístico de reconhecimento baseado em DWT e Modelos Ocultos de Markov (*Hidden Markov Models*, HMM). Primeiro são definidas sub-imagens de mesmo tamanho e com sobreposições, obtidas a partir da imagem original; a seguir se aplica a DWT sobre cada sub-imagem, gerando os vetores de características a partir da magnitude decrescente dos coeficientes DWT; depois, treina-se um modelo HMM para cada face de treinamento, considerando os vetores de características gerados pela DWT. A

tomada de decisão é realizada através da classificação por probabilidade máxima sobre os modelos HMM treinados. O método foi testado sobre o banco de faces ORL, considerando 5 poses de treinamento e 5 de teste e obteve de 97,4% a 100% de acerto. A taxa máxima foi obtida considerando sub-imagens de dimensão 16 x 16, vetores de características com 12 elementos e sobreposição de 50%. Este método é muito eficaz, porém possui um elevado custo computacional.

Jones e Viola [34] propuseram um método não holístico baseado em atributos de Haar e imagem integral (*integral image*). A imagem integral é uma matriz onde cada elemento possui o valor do elemento à esquerda mais o elemento a cima. Está técnica será mais bem detalhada na seção 3.5.2 onde o método de detecção de face de Viola e Jones [67] será explicado. É utilizado o algoritmo AdaBoost para o treinamento, o qual determina os atributos mais relevantes. Esses atributos são utilizados como filtros e aplicados sobre regiões das faces. Esses filtros são analisados em cascata, ou seja, a análise negativa em um desses filtros implica a não análise dos demais. Esta análise em cascata proporciona um menor custo computacional. Aplicado sobro o banco de dados FERET foi atingido uma taxa de reconhecimento de 94%.

Métodos de reconhecimento de faces 3D possuem a vantagem de capturar toda a geometria da face, podendo visualizar detalhes diferenciais como curvatura, profundidade, textura e volume, informações inacessíveis aos métodos 2D. Entretanto, a análise 3D é bem mais complexa e exige um elevado custo computacional. O trabalho de Ansari e Abdel-Mottaleb [3] apresenta um método de reconhecimento de faces onde coordenadas 3D de alguns pontos da face são selecionadas automaticamente. A partir desses pontos é realizado o *morphing* da face para um modelo padrão, normalizando escala e rotação, mantendo a face frontal. Em seguida é realizado o cálculo das distâncias Euclidiana entres esses pontos para determinar a quem pertence a face. Este método obteve taxas de reconhecimento de 96,2% sobre um banco de dados privado com 26 pessoas com dois pares de imagens. A maioria dos outros métodos de reconhecimento de face 3D também utiliza a estratégia de definir pontos principais e depois efetuar o *morphing* para um modelo 3D padrão. O que varia de método para método é a forma de comparar o modelo padrão com as faces obtidas após a aplicação do *morphing* [1].

Um método de reconhecimento de faces não holístico baseado em redes neurais foi proposto por Lin *et al.* [42]. O método utiliza rede neural baseada em decisão probabilística (*Probabilistic Decision-Based Neural Network*, PDBNN). A partir da face de entrada é

extraída uma sub-região contendo apenas os atributos sobrancelhas, olhos e nariz (sem a boca). Essa região é reduzida para o tamanho de 14x10 pixels, e geram-se duas imagens: uma com os pixels normalizados e outra contendo a detecção de bordas da região. Essas imagens alimentam dois PDBNNs e o resultado do reconhecimento é a fusão das saídas dos PDBNNs. Este método atingiu taxas de reconhecimento de 96% sobre ambos os bancos de dados ORL e FERET.

3.5. Métodos de Detecção de Face

Atualmente existem métodos de detecção de face que utilizam diversas abordagens. A seguir é realizado um estudo sobre algumas das mais relevantes, como aquelas baseadas em redes neurais, atributos de Haar, pele humana, e transformadas.

3.5.1. Métodos Baseados em Redes Neurais

Rowley *et al*. [57] descreve um método baseado em redes neurais para detectar faces frontais em imagens monocromáticas. Nele, uma rede neural é treinada com um banco de faces e um banco de "não-faces". O banco de faces é gerado extraindo-se manualmente as faces a partir de imagens obtidas na Internet. Em seguida, a posição de seus atributos (olhos, nariz e boca, entre outros) é definida. Sobre essas imagens são aplicados filtros de rotação, translação e espelhamento, fazendo com que o classificador fique robusto a pequenas variações. O banco de "não-faces" é gerado durante o treinamento, a partir dos falsos positivos encontrados na saída do classificador.

A rede neural recebe como entrada um bloco de imagem de tamanho 20x20 pixels, e retorna um valor entre -1 e +1, onde -1 representa ausência de face, e +1 representa presença de face. Cada imagem a ser classificada é analisada em várias escalas diferentes; cada escala é 1,2 vezes menor que a anterior. Para cada escala uma janela deslizante percorre toda a imagem. De cada posição da janela é extraído um bloco (20x20), ao qual são aplicados filtros de pré-processamento, os quais realizam correção de iluminação, aplicação de máscara de exclusão de background, e equalização de histograma. Esses filtros foram propostos inicialmente por Sung e Poggio [63]. Após a aplicação dos filtros, o bloco é avaliado pela rede neural, utilizando campos receptivos; estes podem ser de três tipos: de tamanho 10x10, 5x5 ou 20x5. Esses campos, mostrados na Figura 7, estão conectados às unidades ocultas da

rede neural, e permitem a localização de atributos de faces humanas, como par de olhos, boca e nariz. Sobre o conjunto de todos os blocos considerados faces, é realizado um pós-processamento, onde blocos isolados são ignorados, e blocos sobrepostos são combinados para formarem um só.

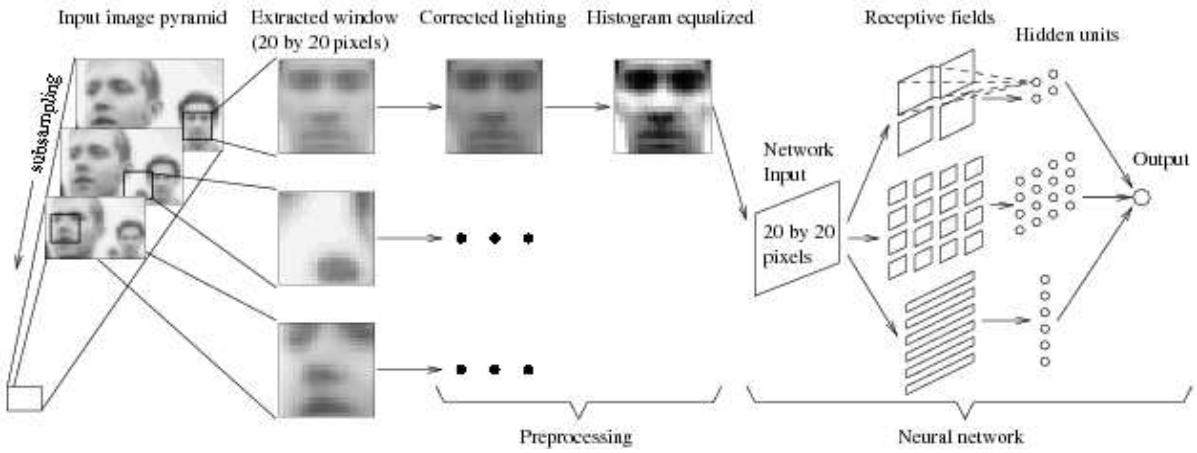


Figura 7: Esquema utilizado por Rowley *et al.* [57].

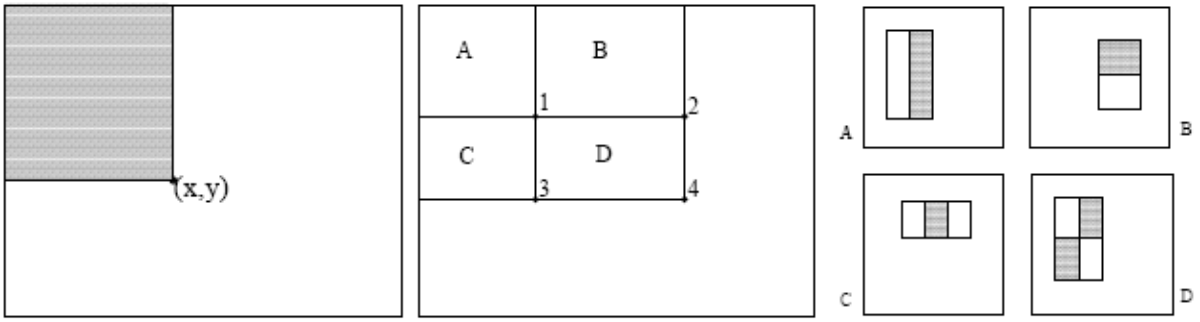
Este método obteve taxas de acerto entre 77,9% e 90,3% sobre o banco de imagens CMU [57]. Entretanto, o tempo de execução desse método não permite que ele seja aplicado sobre vídeo em tempo real.

O trabalho de Rowley *et al*. [57] tornou-se relevante devido ao fato de ter sido um dos primeiros a atingir taxas de acerto na faixa de 90%, e por ter uma didática completa explicando detalhadamente as etapas de pré-processamento, pós-processamento, e obtenção de amostras, além de ter proposto um banco de testes, CMU, largamente utilizado nos trabalhos de detecção de face posteriores a ele [22][27][33][41][47][48][60][67][70][71].

3.5.2. Métodos Baseados em Atributos de Haar

Viola e Jones [67] descrevem um método para detectar faces frontais em imagens monocromáticas. Este foi o primeiro método de detecção de face em tempo real em vídeo, conseguindo processar até 15 quadros por segundo.

Esse método utiliza uma imagem integral para analisar e localizar atributos retangulares rapidamente. A imagem integral é uma matriz do tamanho da imagem a ser analisada, onde cada elemento da matriz possui a soma de todos os níveis de cinza dos pixels à esquerda e acima do pixel atual. A Figura 8.a ilustra uma imagem integral, onde o ponto x,y possui a soma de todos os níveis de cinza da região cinza.



(a) Ponto (x,y) em uma imagem integral.

(b) Pontos 1, 2, 3 e 4 em uma imagem integral, definindo regiões A, B, C e D.

(c) Atributos retangulares em uma imagem integral.

Figura 8: Imagem integral [67].

Na imagem integral, a soma dos níveis de cinza de qualquer região retangular, de qualquer tamanho, pode ser processada com apenas quatro acessos a memória. Por exemplo, considere-se que I_1 representa o valor da imagem integral no ponto 1 indicado na Figura 8.b, I_2 o valor no ponto 2, e assim sucessivamente. A soma dos níveis de cinza da região D pode ser calculada rapidamente como $I_4+I_1-(I_2+I_3)$. A Figura 8.c mostra atributos retangulares de Haar [52][67], o valor desses atributos é calculado como a soma dos pixels da região branca subtraída da soma da região cinza. Utilizando a imagem integral, esses atributos podem ser calculados rapidamente [67].

O método de Viola e Jones [67] utiliza como entrada faces de tamanho 24x24 pixels, e cada imagem a ser classificada é analisada sobre escalas diferentes, cada uma 1,25 vezes menor que a anterior, totalizando 12 escalas diferentes. Esse modelo é similar ao proposto por Rowley *et al*. [57]. Durante o treinamento, os atributos retangulares são localizados e analisados, verificando se são úteis ao classificador. Para isto é utilizado uma variação da técnica *AdaBoost* [20][67], que combina classificadores fracos para formar um classificador forte. A Figura 9 mostra os dois melhores atributos encontrados durante o treinamento. No total, 200 atributos são selecionados e utilizados em cascata [67].

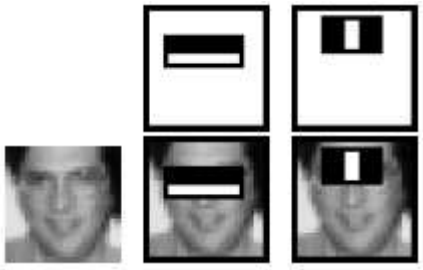


Figura 9: Atributos retangulares de Haar [67].

O algoritmo obteve taxas de acerto entre 76,1% e 93,9% nas 130 imagens do banco de imagens CMU [57][57], com um número de falsos positivos compatível aos de Rowley *et al*. [57]. Esse método está sendo largamente utilizado nos detectores de faces mais recentes, e diversos artigos [33][41][47][48][56][60][70] o utilizam como base. Existe uma implementação de código aberto desse método na biblioteca OpenCV [51].

3.5.3. Métodos Baseados em Regiões de Pele Humana

A pele humana, independentemente de suas variações, tende a seguir uma mesma distribuição probabilística em relação às suas componentes de cor a compõe [68]. Portanto, regiões de pele podem ser definidas e utilizadas como etapa inicial para detectar a presença de faces em uma imagem. Uma limitação deste método é que é necessário que a imagem a ser classificada seja colorida.

As técnicas de detecção de pele podem ser utilizadas em diversos sistemas de cores, como o YIQ, YUV, HSL, RGB, entre outros [66]. O sistema RGB possui três componentes com informações de cor e iluminação; portanto, as três componentes são sensíveis às variações de iluminação [59]. Já os sistemas que possuem uma componente separada para iluminação, e duas componentes de cor, são mais robustos, pois se ignora a componente relacionada à iluminação, observando-se apenas as componentes com informação de cor [66]. Para delimitar as regiões de pele pode-se utilizar uma distribuição Gaussiana [64], ou fazer uso de uma série de premissas [59].

Os métodos mais simples de detecção de face através de pele humana utilizam apenas premissas, encontradas empiricamente, para detectar as regiões de face. Em Buhiyan *et al.* [10], o sistema YIQ é utilizado, e as premissas da Equação (6) determinam as regiões de pele. Em Kovac [40], o sistema RGB é utilizado seguindo as regras da Equação (7). Mesmo simples, esses métodos conseguem um boa taxa de detecção chegando a uma taxa de acerto de 90% em Kovac [40], e de 96% em Buhiyan [10]. As imagens utilizadas para a classificação foram selecionadas especificamente para esses testes, e o banco utilizado não está disponível na Internet para comparações com outros métodos.

$$(60 < Y < 200) \;\&\; (20 < I < 50) \tag{6}$$

$$\begin{aligned} &((\max(R,G,B) - \min(R,G,B) > 15) \;\&\; \\ &(R > 95) \;\&\; (G > 40) \;\&\; (B > 20) \;\&\; \\ &(|R-G| > 15) \;\&\; (R > G) \;\&\; (R > B)) \\ &\text{OR} \\ &((R > 220) \;\&\; (G > 210) \;\&\; (B > 170) \;\&\; \\ &(|R-G| \leq 15) \;\&\; (R > B) \;\&\; (G > B)) \end{aligned} \tag{7}$$

Os métodos, que fazem uso da distribuição Gaussiana utilizam sistemas de cores que tenham uma componente com informação de iluminação e duas componentes com informação de cor. A componente de iluminação é ignorada e verifica-se como é a distribuição das duas componentes de cor em um plano 2D. A partir dessa distribuição ajusta-se- um modelo Gaussiano à distribuição, e determinam-se seus limiares. Em seguida é verificado se cada pixel da imagem a ser classificada está contido na região útil desse modelo. A Figura 10 apresenta um exemplo de histograma 2D, em escala logarítmica, das componentes I e Q do sistema de cores YIQ para imagens de pele humana [59].

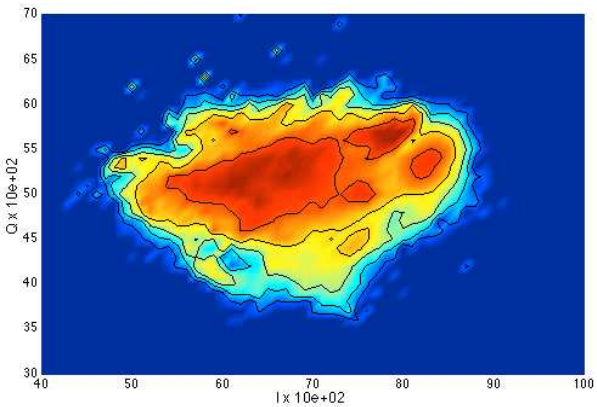


Figura 10: Histograma 2D, em escala logarítmica, das componentes I e Q dos sistema de cores YIQ para imagens de pele humana, Terrilon *et al.* [64].

Diversos métodos de reconhecimento de face utilizam como etapa inicial o reconhecimento de regiões de pele humana [2][5], e às vezes como etapa única [10][40]. Esses métodos realizam a detecção em tempo real.

3.5.4. Métodos Baseados em Transformadas

Na área de análise de faces, o trabalho com transformadas teve início no reconhecimento de faces, contudo ela passou a ser utilizada também na detecção de faces.

Popovici e Thiran [54] propuseram um método de detecção de faces que utiliza PCA para gerar um modelo de faces em um espaço de faces (*Eigenfaces*). Na classificação é utilizado o algoritmo *Support Vector Machines* (SVM). A SVM foi treinada de duas formas, uma utilizando um núcleo polinomial e outra utilizando um núcleo de funções radiais. Este método obteve uma taxa de detecção de 97,93% sobre o banco de dados BANCA utilizando um espaço *Eigenfaces* com 102 dimensões.

O trabalho de Kervrann *et al.* [37] propõe um detector de faces baseado em *Eigenfaces*. Este método define dois modelos, um de faces e outro de não faces, gerados a partir do espaço *Eigenfaces*. É utilizado um critério de casamento baseado na Razão de Verossimilhança Generalizada (*Generalized Likelihood Ratio*, GLR), e é realizada uma otimização através do algoritmo de arrefecimento simulado (*simulated annealing*). Os testes desse algoritmo são realizados em um banco de imagens particular e não foram divulgadas taxas de acerto.

Capítulo 4

Materiais e Métodos

No presente trabalho elaborou-se uma arquitetura para um sistema de detecção e reconhecimento de face em vídeo em tempo real. O sistema foi implementado e mostrou-se eficiente atingindo taxas de detecção e reconhecimento compatíveis com o estado da arte. Neste capítulo são descritos os materiais e métodos utilizados no desenvolvimento deste trabalho.

4.1. Ambiente de Desenvolvimento

O sistema de detecção e reconhecimento facial proposto foi desenvolvido em um computador com processador de 2,0 GHz, AMD Turion64 de núcleo único e com 1GB de RAM, executando o Ubuntu Linux. O sistema completo está desenvolvido em C++, e o módulo de reconhecimento possui duas versões, uma C++ e outra Java.

4.2. Bancos de Faces

Existem diversos bancos de dados de faces utilizados para comparações entre os vários métodos de reconhecimento e detecção de face. Alguns, como o CMU e MIT são mais utilizados por métodos de detecção, outros, como o ORL, são mais utilizados na validação de métodos de reconhecimento. Neste trabalho, além dos bancos padrão disponíveis na Internet,descritos a seguir, foi criado um banco de dados próprio para validar a arquitetura do sistema detector/reconhecedor, batizado como Banco de Faces UFPB.

4.2.1. Banco de Faces do *Massachusetts Institute of Technology*, MIT

O banco de dados do *Massachusetts Institute of Technology*, MIT, foi utilizado inicialmente por Sung e Poggio [63] e contém 23 imagens, as quais possuem um total de 155 faces em escala de cinza e em baixa resolução. Essas imagens ilustram cenas variadas e possuem faces frontais com pequenas variações de ângulo [63]. A Figura 11 ilustra algumas imagens desse banco de dados.

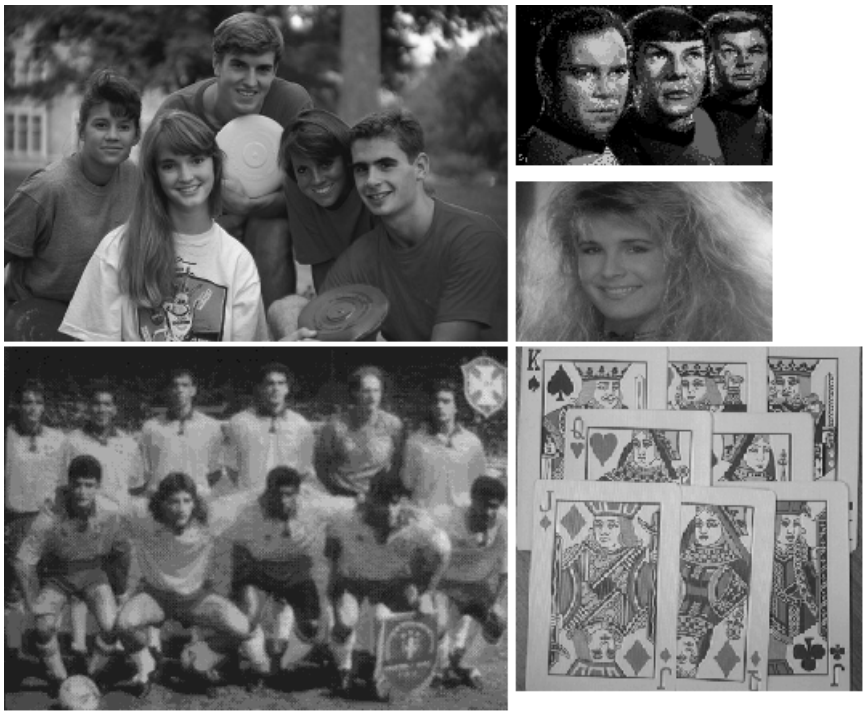


Figura 11: Exemplos de imagens do banco de dados MIT.

Este banco de dados é mais utilizado em associação com o banco de dados CMU.

4.2.2. Banco de Faces da *The Carnegie Mellon University*, CMU

O banco de faces da *Carnegie Mellon University*, CMU, foi utilizado inicialmente por Rowley *et al*. [57] para testar seu método de detecção de face. Este banco de dados possui 130 imagens, totalizando 507 faces frontais, algumas rotacionadas. Entre essas 130 imagens estão as 23 imagens do banco de dados MIT. Por conta deste fato esse banco de dados também é conhecido com CMU/MIT. Este banco de dados possui fotos em escala de cinza, em vários tamanhos, em diversas cenas e com diferentes e complexos backgrounds, e as

pessoas encontram-se em diversas situações, como caminhando, conversando, sentadas, em pé, etc [57]. A Figura 12 exemplifica algumas imagens desse banco de dados.

Figura 12: Exemplo de imagens do banco de dados CMU.

Este banco de dados passou a ser largamente utilizado, como base de comparação, em trabalhos de detecção de face posteriores [22][27][33][41][47][48][60][67][70][71] ao trabalho de Rowley *et al*. [Ref]. No presente trabalho, este banco de dados foi utilizado para validar o detector de faces utilizado.

4.2.3. Banco de Faces *Olivetti Research Lab,* ORL

O Banco de faces ORL (*Olivetti Research Lab*) foi desenvolvido nos laboratórios da Olivetti em Cambridge, Inglaterra [4]. Este banco de dados é composto de 400 fotos de faces de 40 pessoas diferentes. Cada pessoa possui 10 fotos em diferentes poses, com pequenas variações de iluminação, expressões faciais, e acessórios. Todas as imagens foram fotografadas contra um fundo homogêneo, e as faces estão todas verticais e frontais, com pequenas variações de angulação. As imagens são todas em escala de cinza com 256 níveis de cinza, e possuem tamanho de 92x112 pixels [4]. A Figura 13 mostra algumas faces do banco ORL.

Figura 13: Banco de dado ORL: pessoas 1, 10, 20 e 35, em suas 10 poses.

Este banco de faces é muito utilizado para validar métodos de reconhecimento de face [9][26][39][45], e também para o treinamento de métodos de detecção [19][38][49].

No presente trabalho, a etapa de reconhecimento de face foi treinada e testada utilizando o banco de dados ORL.

4.2.4. Banco de Faces UFPB

O banco de faces UFPB foi criado para validar a arquitetura de detecção e reconhecimento de face proposto neste trabalho. Foi necessária sua criação porque os bancos de faces de vídeo, reportados na literatura, possuem apenas vídeos individuais das pessoas, não havendo um vídeo onde a classificação pudesse ser avaliada sobre múltiplas pessoas.

Este banco foi gerado em dois dias, a partir de 40 pessoas, entre alunos, professores e funcionários da UFPB que trabalham nos laboratórios de pesquisa Lavid (Laboratório de Aplicações de Vídeo Digital) e Lasid (Laboratório de Sistemas Distribuídos). Este banco possui duas versões: uma em vídeo, e outra em fotos. Todas as faces capturadas são frontais, com pouca rotação.

O banco de faces UFPB-Vídeo, possui um vídeo de aproximadamente 10 segundos para cada pessoa. Os vídeos são coloridos no formato YUV, possuem resolução máxima de 1920x1080 pixels, e foram capturados a 29,7 quadros por segundo, a partir de uma câmera de vídeo de alta definição (*full* HD), armazenados no formato *mpeg2* [29], sem áudio. Nos vídeos, cada pessoa executa pequenos movimentos circulares com a cabeça e varia um pouco as expressões faciais. A Figura 14 ilustra esses movimentos em duas pessoas do banco de dados. Essas variações permitem que sejam capturadas uma maior variedade de poses. Este

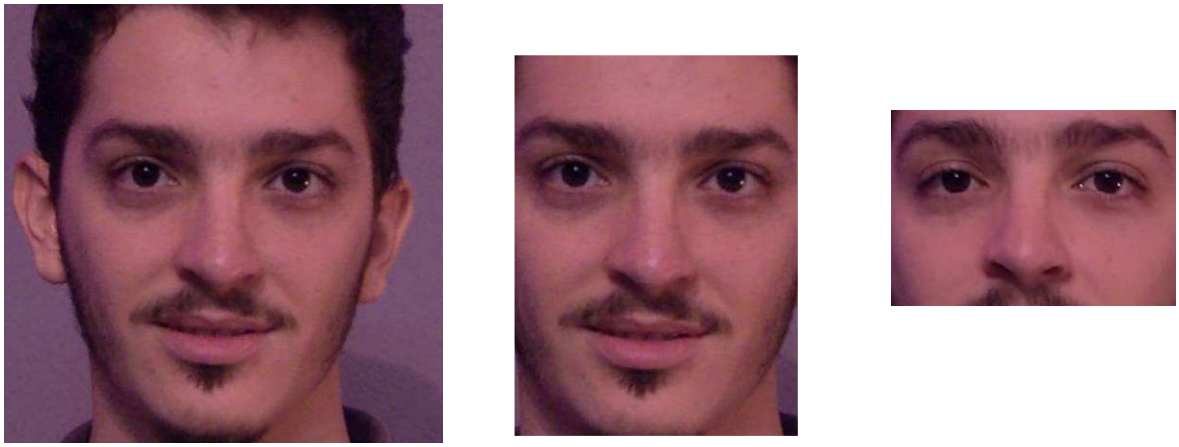
banco de dados ainda possui um vídeo coletivo que exibe 22 pessoas, das 40 pessoas do banco completo, executando suas tarefas cotidianas nos laboratórios. Esse vídeo serve para validar o reconhecedor de faces. Um quadro desse vídeo é ilustrado na Figura 15.

Figura 14: Exemplos de imagens do banco de dados UFPB-Vídeo.

Figura 15: Exemplo de quadro do vídeo de classificação do banco de dados UFPB-Vídeo.

O banco de faces UFPB-Fotos foi gerado a partir da aplicação do módulo detector de face sobre o banco de dados UFPB-Vídeo. Centenas de faces foram detectadas para cada pessoa. Foi então realizada uma aglomeração por (*clustering*) por k-médias (*k-means*) [43] para selecionar as 20 mais representativas. O algoritmo k-médias agrupa um conjunto de *n* amostras em *k* clusters, de acordo com alguma métrica de similaridade do espaço de atributos [43].

A partir das imagens selecionadas, o banco de faces UFPB-Fotos foi dividido em três grupos, onde cada grupo possui 20 poses por pessoa. O primeiro (UFPB-Fotos-1) contém as faces completas na forma como foram detectadas. O segundo (UFPB-Fotos-2) é uma sub-região retangular central do primeiro, a qual exclui parte do cabelo e do background. O terceiro (UFPB-Fotos-3) é uma sub-região do segundo contendo apenas a região que engloba sobrancelhas, olhos, e nariz, excluindo a boca. A Figura 16 ilustra estes grupos. No subgrupo UFPB-Fotos-2, a redução da área permite a redução de background. Já no subgrupo UFPB-Fotos-3 esta redução permite uma relativa invariabilidade referente a expressões faciais, ao corte/penteado/movimento do cabelo, e a movimentos bucais. O banco de faces UFPB-Fotos-3 foi inspirado nos trabalhos de Lin *et al.* [42] e de Sorobich e Kirby [62], os quais também utilizaram regiões reduzidas similares. As imagens finais são em cores (RGB) e estão armazenadas no formato *Bitmap* (BMP), que não emprega compressão com perdas. O subgrupo UFPB-Fotos-1 possui imagens de tamanho 512x512 pixels, o subgrupo UFPB-Fotos-2 de 320x432 pixels, e o subgrupo UFPB-Fotos-3 de 320x218 pixels.



(a) Face completa - UFPB-Fotos-1 **(b) Sub-região 2 - UFPB-Fotos-2** **(c) Sub-região 3 - UFPB-Fotos-3**

Figura 16: Grupos de imagens presentes no banco de faces UFPB-Fotos.

O banco de faces UFPB completo está disponível no sítio http://www.lavid.ufpb.br/~ufpbdatabase .

4.3. *Open Computer Vision Library*, OpenCV

A *Open Computer Vision Library* (OpenCV) é uma biblioteca multiplataforma para o desenvolvimento de aplicativos na área de visão computacional. Foi originalmente desenvolvida pela Intel, entretanto é totalmente livre ao uso acadêmico e comercial, desde que siga a licença BSD da Intel. A maior parte de suas funções é implementada em C/C++, porém há suporte a outras linguagens. Esta biblioteca possui módulos de processamento de imagens e vídeo, interface gráfica básica, controle de periféricos (mouse/teclado), além de diversos algoritmos de visão computacional, como segmentação de imagens, detecção de objetos, detecção de movimento e outros diversos filtros [11][51].

Essa biblioteca é utilizada neste trabalho nos módulos de captura de vídeo, detecção de face, e no módulo de exibição.

4.4. Arquitetura do Sistema Proposto

O presente trabalho apresenta uma arquitetura para um Sistema de Detecção e Reconhecimento de Faces, SDRF. Esse sistema recebe como entrada um vídeo com faces para classificação, e tem como saída o vídeo com o resultado da detecção e do reconhecimento das faces presentes.

A arquitetura proposta é composta por quatro módulos: módulo de captura de vídeo, módulo detector de face, módulo reconhecedor de face, e módulo de exibição. O módulo de captura realiza a obtenção das imagens de entrada; elas podem ser adquiridas de um arquivo de vídeo previamente capturado, ou pode ser lida diretamente de algum dispositivo de captura de vídeo. Neste módulo, o vídeo é decodificado e transformado em uma seqüência de quadros, os quais serão passados ao módulo de detecção de face.

O módulo de detecção de face analisa todos os quadros de vídeo e verifica a existência ou não de faces na cena, indicando a posição espacial onde as faces se encontram. O módulo de reconhecimento de face analisa apenas as regiões que foram consideradas como face pelo detector de face. Sobre essas regiões é então realizado o reconhecimento de face,

que indica a que pessoa pertence a face em análise. O módulo de exibição simplesmente exibe o resultado da detecção e reconhecimento sobre o vídeo em análise. A Figura 17 ilustra o funcionamento do sistema completo.

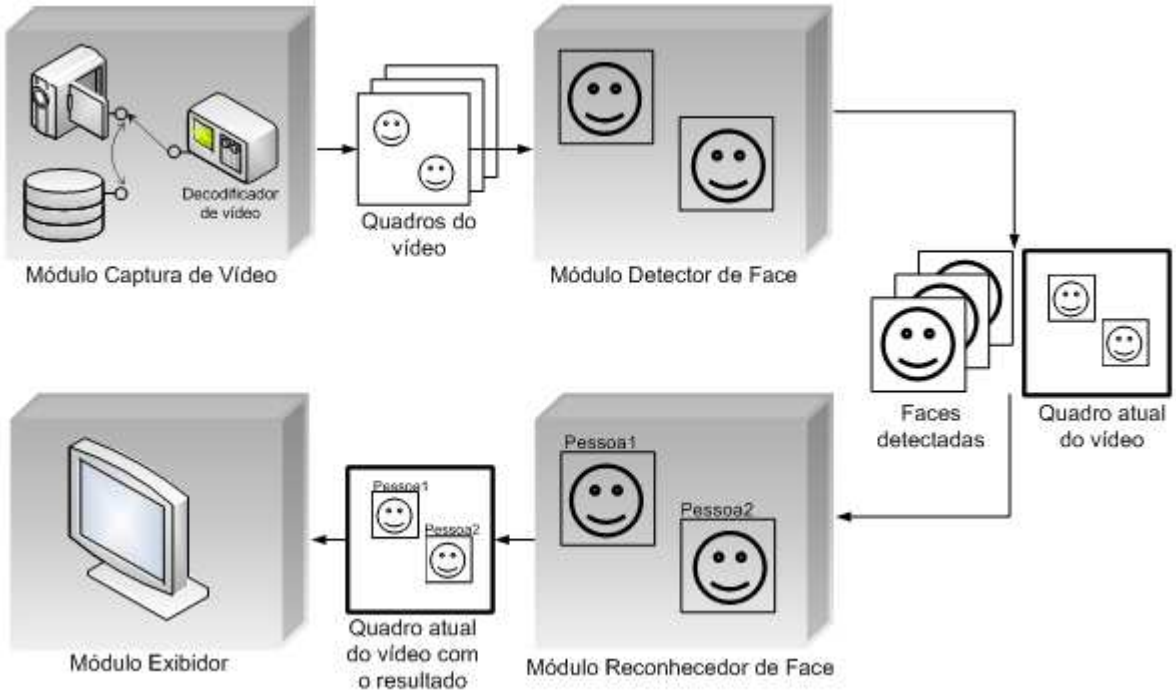


Figura 17: Arquitetura geral do SDRF.

Os módulos de captura e exibição de vídeo são a entrada e a saída do sistema respectivamente, e não necessitam de detalhamento. Já os módulos de detecção e reconhecimento serão detalhadamente descritos a seguir.

4.5. Módulo Detector de Face

Inicialmente tentou-se desenvolver uma nova técnica de detecção de face, inexistente entre os trabalhos atualmente encontrados na literatura. Para tal, vários experimentos foram realizados utilizando-se diversas técnicas. Entretanto, nenhuma dessas técnicas obteve resultados compatíveis com as publicações mais recentes de detecção de face, por isso a versão final do módulo detector utiliza o método de Viola e Jones [67], que é consagrado, e está implementado em código aberto, disponível na biblioteca OpenCV [51]. As tentativas de desenvolvimento de novos métodos de detecção de face são descritas na seção 4.5.2.

4.5.1. Método Utilizado

O detector de faces utilizado neste trabalho é uma adaptação do detector de faces do OpenCV, que por sua vez é baseado no método de Viola e Jones [67]. Este detector possui uma implementação em C, e foi realizado um encapsulamento em classes C++ para uma melhor integração ao sistema.

A biblioteca OpenCV possui funções para o treinamento do detector de faces, e também possui exemplares de resultados do treinamento previamente armazenados. Esses resultados são armazenados na forma de arquivos XML, e podem ser utilizados diretamente sem necessidade de realizar um novo treinamento. Neste trabalho não está sendo realizado um novo treinamento, e o resultado de um treinamento previamente realizado pela OpenCV é apenas carregado. O arquivo XML com o treinamento utilizado foi o *haarcascade_frontalface_alt2.xml*.

O detector possui três parâmetros de configuração, o tamanho mínimo de faces a ser buscado, razão de zoom entre as diferentes escalas da imagem a ser classificada, e a sub-região a ser extraída. O detector foi utilizado com o tamanho mínino de faces igual a 20x20 pixels, e com a razão de zoom igual a 1.2. A sub-região extraída é utilizada para determinar qual subárea da face vai ser utilizada no Módulo de Reconhecimento. Estão sendo utilizadas três sub-regiões padrões visando uma compatibilidade com as três regiões de faces do banco de dados UFPB-Fotos. A primeira região é a face inteira, sendo compatível com as imagens do banco de faces UFPB-Fotos-1. A segunda região é uma sub-região da primeira, contendo a área equivalente a do banco de faces UFPB-Fotos-2. A terceira região é uma sub-região da segunda, equivalente ao banco UFPB-Fotos-3.

Todo o processo de detecção é realizado sobre imagens em escala de cinza. Imagens são coloridas são convertidas para escala de cinza, para a detecção.

4.5.2. Outros Métodos Avaliados

Foram realizadas experimentações com as técnicas da Entropia, Predição por Casamento Parcial (*Prediction by Partial Match*, PPM) e DCT. Todas essas técnicas foram aplicadas seguindo basicamente uma mesma abordagem.

Foi gerado um banco de dados de faces de tamanho 20x20 para treinamento. Esse tamanho foi utilizado inspirado em diversos trabalhos da literatura [57][67] que utilizam esse

tamanho ou algo bem próximo. A partir desse banco, gera-se um modelo de entropia, PPM ou DCT das faces.

Sobre a imagem a ser analisada percorre-se uma janela deslizante de tamanho 20x20. A região correspondente é comparada com o modelo construído no aprendizado, e decide-se se a região contém ou não uma face. Para permitir que faces de tamanho maior que 20x20 também sejam detectadas, o procedimento é repetido sobre várias versões em escalas reduzidas da imagem original.

4.5.2.1. Entropia

A entropia mede a quantidade de informação presente em uma determinada mensagem, ou imagem. No experimento realizado neste trabalho, o modelo de entropia de uma face é composto de um intervalo, com o valor máximo e mínimo permitidos para a entropia de uma face. Este intervalo é obtido a partir da análise do conjunto de valores de entropia das faces de determinado banco de faces. Deste conjunto são extraídos o maior e o menor valor para delimitar o intervalo de entropias permitido.

Quando em modo de classificação, sobre cada bloco extraído da janela deslizante é realizado o cálculo de sua entropia. Se a entropia estiver dentro do intervalo do modelo de faces então o bloco é considerado como uma possível face, senão é desprezado. Apenas a entropia não é suficiente para determinar se um bloco é face ou não, então ela sempre é utilizada em conjunto com o PPM ou com a DCT, funcionando como um filtro de exclusão.

4.5.2.2. PPM

O PPM contém uma técnica de modelagem estatística muito eficiente. Atualmente os compressores com as melhores taxas de compressão de dados sem perdas utilizam alguma versão do PPM. Devido a sua capacidade de construir rapidamente modelos estatísticos precisos, o PPM também vem sendo utilizado na área de reconhecimento de padrões [28][44].

Foram realizados testes para verificar a aplicabilidade do PPM para detecção de face. Na abordagem utilizada, o universo de imagens foi dividido em duas classes: faces e não-faces. Para cada classe foi gerado um modelo probabilístico da distribuição dos pixels

utilizando o PPM. Ambas os modelos foram gerados a partir de diversas imagens de tamanho 20x20.

Sobre o bloco extraído da janela deslizante é aplicado o filtro da Entropia. Se este filtro não descartar o bloco, realiza-se o cálculo da razão de compressão sobre o modelo de faces e sobre o modelo de não-faces. Se a razão de compressão for maior com o modelo de faces, considera-se que o bloco em análise é uma face; se for melhor com o modelo de não-faces então o bloco não é uma face.

Essa metodologia, contudo, não se mostrou eficiente, visto que os pixels em uma face variam muito, e são muito sensíveis a alterações de iluminação, escala e rotação, e o PPM não se adapta bem a essas variações.

4.5.2.3. DCT

A DCT é uma transformada que tem a propriedade de concentrar grande parte da energia de um sinal em uma pequena região. Essa propriedade é explorada no sentido de selecionar os atributos mais relevantes de uma face.

O modelo de faces é gerado convertendo-se as imagens do banco de dados de treinamento para suas respectivas DCTs. Em seguida é realizada a seleção dos atributos mais relevantes, através de um seletor de baixas freqüências. Esse seletor mantém os coeficientes de menor freqüência da DCT e despreza os demais. Por fim, é realizado o cálculo da distância de cada uma das DCTs do banco de treinamento para todas as demais DCTs de treinamento. Desse cálculo é obtido o valor da maior distância permitida para que uma DCT seja considerada como pertencente ao modelo de faces. A distância mínima sempre será zero.

No modo de classificação, cada bloco da janela deslizante é extraído, e então é aplicado o filtro da Entropia. Se este filtro não descartar o bloco, gera-se sua DCT e selecionam-se os atributos. Em seguida é calculada a distância da DCT do bloco de teste para todas as DCTs do modelo de faces. Se for encontrada alguma distância inferior à distância máxima permitida então esse bloco é considerado face, senão é considerado não face.

Contudo, esse algoritmo não se mostrou eficiente para detecção de face. Uma possível explicação para este fato é que a DCT não conseguiu generalizar as características específicas de cada pessoa para a construção de um modelo único de faces. Entretanto, a DCT mostra-se

eficiente para reconhecimento de face, conforme será demonstrado no módulo de reconhecimento de face.

4.6. Módulo Reconhecedor de Face

Neste trabalho foi desenvolvido um novo método holístico de reconhecimento de face baseado na distância entre os coeficientes das DCTs. Todas as imagens de face a serem analisadas e as pertencentes ao banco de faces são convertidas para o domínio da freqüência gerando-se sua DCT. Os coeficientes mais relevantes para o reconhecimento são selecionados utilizando a técnica de seleção de baixas freqüências [45]. Durante a classificação de uma face é calculada a sua DCT e são selecionados seus coeficientes. Em seguida é calculada a distância entre os coeficientes selecionados da DCT da face a ser classificada e os coeficientes selecionados das DCTs de todas as faces do banco de dados. A menor de todas as distâncias provavelmente ocorrerá entre faces da mesma pessoa. Por isso, a face em análise será classificada como pertencente à pessoa na qual ocorreu a menor distância. Esse método é aplicado sobre imagens em escala de cinza; quando as imagens estiverem em cores têm de ser convertidas.

A técnica da distância de Minkowski de ordem um foi adaptada e utilizada para o cálculo das distâncias entre os coeficientes DCT de face. Essa distância foi calculada como a soma dos módulos da diferença entre os valores absolutos dos coeficientes DCT, como especificado pela Equação (9). A adaptação realizada foi a utilização dos valores absolutos dos coeficientes. Na distância de Minkowski de ordem um original, Equação (8), não são utilizados os valores absolutos. Este cálculo é simples, pois não envolve computação de raízes nem quadrados, tendo um baixo custo computacional.

$$\sum_{i=1}^{n} \left| x_i - y_i \right| \qquad \textbf{(8)}$$

$$\sum_{i=1}^{n} \left\| x_i \right| - \left| y_i \right\| \qquad \textbf{(9)}$$

Alguns métodos descritos na literatura também são baseados em transformadas matemáticas, como DCT, transformada de *Karhunen-Loève*, e transformada de Wavelet [45]. Alguns deles são baseados em DCT, e a utilizam por causa de suas propriedades de

concentração de energia e devido à existência de algoritmos rápidos para sua computação. Os métodos baseados em DCT atingem taxas de acerto compatíveis com os baseados em KLT e Wavelet [17].

Determinados métodos baseados em DCT não são holísticos, e realizam o cálculo da DCT em regiões específicas da imagem, como olhos, nariz e boca [9][39][53]. Outros métodos são holísticos, e utilizam a DCT para realizar a seleção dos atributos, e então utilizam classificadores baseados em similaridades, como o K-Vizinho Mais Próximo (KNN) ou Distância Mínina ao Protótipo [26][45].

Utilizando DCT, Hafed e Levine [26] atingiram taxas de reconhecimento de 93%. Podilchuk e Xiaoyu [53] obtiveram um taxa de acerto de 94%, enquanto Matos *et al.* [46] obteve 99,25%. Apesar de Kumar *et al.* [39] e Bicego *et al.* [9] atingirem 100% de acerto, eles utilizam a DCT em conjunto com Modelos Ocultos de Markov (*Hidden Markov Models*, HMMs), o que leva a um sistema com alto custo computacional.

O método proposto está dividido em três etapas: seleção de atributos, treinamento e classificação. A Figura 18 exibe uma generalização de um sistema de reconhecimento de face. A etapa de seleção de atributos extrai os atributos. A etapa de treinamento seleciona o melhor grupo de atributos para realizar a tarefa de identificação, considerando todas as amostras de treinamento. E a etapa de classificação compara uma face desconhecida com as que estão no banco de faces, utilizando os atributos selecionados para determinar a melhor correspondência.

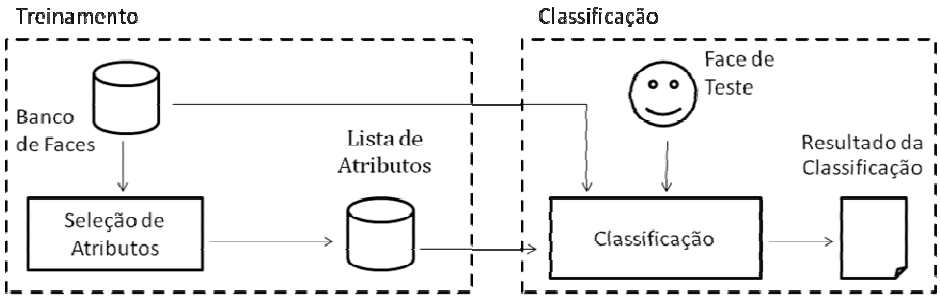


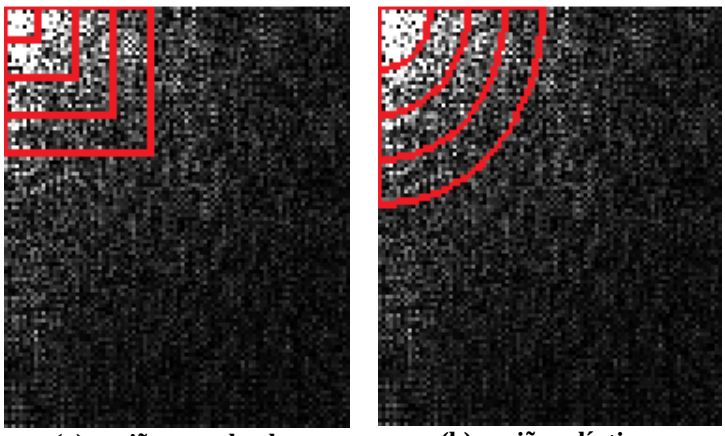
Figura 18: Esquema de reconhecedor de faces, adaptado de Matos [45].

O presente trabalho propõe novas abordagens para selecionar os coeficientes DCT que serão utilizados como atributos das faces, e aplica o classificador do Vizinho Mais Próximo (NN) na etapa de classificação. Este é um aperfeiçoamento aos métodos de Hafed *et al.* [26] e Matos [45].

4.6.1. Seleção de Atributos

A fase de seleção de atributos especifica a lista de características que representarão as faces, considerando as diversas poses presentes no banco de dados de treinamento. O processo de seleção de atributos extrai as informações mais relevantes para distinguir uma pessoa de outra [16]. Desta forma, a dimensionalidade do universo de atributos é reduzida para a quantidade de atributos extraídos.

Para realizar a extração de atributos, a abordagem utilizada é a seleção de coeficientes situados na região de baixas freqüências da DCT. Esta abordagem é rápida e simples, pois não analisa os coeficientes DCT da imagem, nem executa cálculos ou qualquer tipo de comparação. Como a DCT de uma imagem é um sinal 2D, a seleção dos coeficientes de baixa freqüência consiste simplesmente em definir geometricamente uma região no canto superior esquerdo da imagem DCT. Exemplos utilizando regiões quadradas e circulares são apresentados na Figura 19. Esta figura exibe o resultado da aplicação da DCT sobre uma imagem, seguida por uma etapa de normatização para melhor visualização. Como pode ser visto, a maior concentração de energia ocorre na região das baixas freqüências. Por isso, esta abordagem é adequada para captar os coeficientes mais importantes da imagem [45].



(a) regiões quadradas **(b) regiões elípticas**

Figura 19: DCT normalizada da pessoa 4, pose 1, do banco de dados ORL, e as regiões de baixas freqüências.

Um trabalho anterior [46] utilizou regiões quadradas para realizar a extração dos coeficientes, como na Figura 19.a. No presente estudo foram realizados testes com regiões quadradas, retangulares, triangulares e elípticas.

Na Figura 19.b pode ser visto que as regiões elípticas se adaptam melhor à evolução espacial dos coeficientes DCT, selecionando melhor os coeficientes mais importantes, ou seja, de maior amplitude ou energia (pontos mais brilhantes na imagem). Nos testes de

reconhecimento realizados, as regiões de seleção de atributos elípticas mostraram-se superiores às quadradas, obtendo uma maior taxa de acertos.

4.6.2. Treinamento

A etapa de treinamento realiza várias vezes a etapa de seleção de atributos, alterando a cada repetição, as dimensões e a escala das regiões geométricas da seleção. Em seguida, é escolhida a seleção que produz o melhor resultado. Os testes foram realizados utilizando regiões quadradas, retangulares, triangulares e elípticas. Também foram realizados testes com filtros de pré-processamento aplicados às imagens de faces antes de serem convertidas para o domínio da freqüência. Esses testes analisaram a aplicação da equalização e expansão de histograma das imagens. Entretanto, não apresentaram bons resultados e foram descontinuados.

Para determinar qual a melhor região geométrica a ser utilizada na seleção de atributos foi realizado um teste, onde todas as possibilidades de dimensões das formas geométricas utilizadas foram testadas. Para evitar que o teste tenha um alto custo computacional, foi realizada uma redução da área de busca para apenas o primeiro quadrante da DCT das faces. No caso do banco de faces ORL, como as imagens possuem dimensões de 92x112 pixels, foi utilizado apenas o primeiro quadrante da DCT, que possui dimensão de 46x56 pixels. Esta redução da área de busca não teve impacto na obtenção dos melhores resultados, pois a maior parte da energia do sinal se encontra no primeiro quadrante da DCT, como pode ser visto na Figura 20.

Figura 20: Quadrantes da DCT.

Para o banco de faces ORL, o melhor resultado encontrado foi com uma região elíptica de semi-eixos de dimensão 6x7, contendo 29 coeficientes. Para cada banco de dados

diferente, e a cada inserção de nova pessoa no banco de dados, todo o treinamento deve ser, em princípio, refeito, para manter o desempenho ótimo. Isto ocorre porque o desempenho de um classificador depende da inter-relação entre a quantidade de atributos selecionados (dimensão do espaço) e a quantidade de amostras das classes, conforme esclarecido pelo Problema da Dimensionalidade, explicado na seção 2.4.4 (Classificação Estatística).

4.6.3. Classificação

A etapa de classificação compara a face que está sendo classificada com uma série de faces de um banco de dados, considerando apenas os atributos selecionados da DCT, e indica com qual delas a face de teste mais se parece.

Neste trabalho, a comparação é realizada utilizando-se a distância de Minkowski de ordem um, modificada. A abordagem de classificação utilizada é a do Vizinho Mais Próximo (NN). Esta abordagem foi escolhida porque apresentou os melhores resultados em trabalhos anteriores, como o de Matos *et al*. [46].

4.7. Métodos de Avaliação

Nesta seção são apresentadas as técnicas utilizadas para avaliar o funcionamento do SDRF. Os módulos Detector e Reconhecedor são analisados individualmente, e em seguida, o SDRF é analisado por completo.

4.7.1. Métodos de Avaliação do Detector

Os testes realizados sobre o módulo detector de faces foram bastante simples, visto que esse módulo foi desenvolvido baseado no detector da OpenCV, já amplamente conhecido e avaliado pela comunidade de visão computacional.

Para avaliar a eficácia isolada do detector de face, este foi submetido a um teste de classificação sobre o banco de dados CMU/MIT [57], visto que esse banco de dados é largamente utilizado para validar detectores de face.

4.7.2. Métodos de Avaliação do Reconhecedor

Foram realizados quatro testes para validar o método de reconhecimento proposto. O banco de dados ORL foi utilizado em todos os testes e o banco de dados UFPB-Fotos foi utilizado apenas no primeiro. Neste primeiro teste foi aplicada a abordagem de validação cruzada, deixe-um-de-fora. Em seguida, foi analisado o deixe-um-de-fora acumulativo. Após isso, foi realizado um teste de aplicação de zoom manual sobre o único erro encontrado no primeiro teste. Por último, foi realizado um teste utilizando a abordagem deixe-um-de-fora, sendo que a cada face a ser classificada foi aplicado um filtro de zoom automático.

4.7.2.1. Validação Cruzada

A abordagem de teste de validação cruzada, deixe-um-de-fora, mantêm uma face fora do conjunto de treinamento e realiza a classificação com esta face [25]. Como o banco de faces ORL possui dez poses para cada pessoa, dez rodadas são realizadas. Na primeira rodada, a primeira pose de todas as pessoas é excluída do conjunto de treinamento, e são utilizadas como faces de teste a serem classificadas. Na segunda rodada, a segunda pose é excluída do conjunto de treinamento e é utilizada como face de teste, e assim por diante. Esse teste foi realizado diversas vezes, com variações na quantidade de coeficientes, selecionados a partir de diferentes formas geométricas. O banco de dados UFPB também foi utilizada neste teste; a única diferença é que neste banco existem vinte faces para cada pessoa, então vinte rodadas deixe-um-de-fora foram realizadas.

A taxa de reconhecimento é calculada como o total de acertos do reconhecedor dividido pelo total de faces presentes no banco de dados. O resultado desta divisão é então multiplicado por 100, para obtenção das taxas percentuais.

4.7.2.2. Validação Cruzada Acumulativa

O teste de validação cruzada acumulativa mede o reconhecimento acumulativo. Neste teste, o classificador retorna uma lista ordenada, em ordem crescente de distâncias, das faces mais semelhantes à face de teste. Um classificador NN comum iria utilizar apenas o primeiro elemento da lista (o que possui a menor distância) para realizar a classificação. O classificador acumulativo indica o tamanho mínimo da lista para que o classificador atinja 100%. Isto é, o tamanho mínimo da lista que contém uma pose da mesma pessoa da face de

teste. Este teste também foi realizado utilizando diferentes formas geométricas de seleção de atributos.

O reconhecimento acumulativo é indicado, por exemplo, para casos onde o reconhecedor de face funciona em modo de autenticação, e não em modo de identificação. Em modo de autenticação o usuário se identifica inicialmente, através de algum login e senha, e o sistema verifica a autenticidade do usuário através de sua face. Para verificar a autenticidade pode ser utilizado o reconhecimento acumulativo, que verifica se o usuário identificado está presente na lista de faces retornadas.

4.7.2.3. Zoom Manual

Uma análise sobre o erro encontrado no melhor resultado do teste do deixe-um-de-fora sobre o banco de faces ORL, mostrou que, a pessoa 19, em sua pose 9, foi atribuída a pessoa 11, sendo mais parecida com ela na pose 5. A Figura 21 mostra as 10 poses da pessoa 19. Nela, a pose 9 está sublinhada e a última imagem da segunda linha é a pose 5 da pessoa 11. Pode ser visto que a pose 9 da pessoa 19 tem uma face ligeiramente menor do que as suas outras nove poses. Além disso, também pode ser visto que o rosto da pessoa 11 em sua pose 5 tem um tamanho mais próximo ao da pessoa 19 em sua pose 9 do que o resto das poses da pessoa 19.

Figura 21: Banco de faces ORL: Pessoa 19 em suas 10 poses, e a pessoa 11 em sua pose 5.

Foi realizado um teste para verificar se o reconhecedor é realmente sensível à operação de zoom. Neste teste foi aplicado, manualmente, uma ampliação de 5% sobre a pose 9 da pessoa 19. Com esta ampliação, esta pose passa a ter um tamanho mais próximo às outras poses da pessoa de 19, como pode ser visto na Figura 22. Com essa alteração manual, o reconhecedor atingiu 100% de acerto. Isto indica que o reconhecedor baseado em DCT é sensível à operação do zoom. Esse teste foi aplicado apenas sobre o banco de faces ORL, porque o banco de faces UFPB possui variações desprezáveis de escala. Outros valores de

zoom também foram testados, mas com o de 5% atingiu-se 100% e a face ficou visualmente com o tamanho mais próximo ao das demais poses da pessoa 19.

Figura 22: Banco de faces ORL: Pessoa 19 em suas 10 poses. A pose 9 esta aumentada em 5%.

4.7.2.4. Zoom Automático

Com o objetivo de tornar a classificação robusta a pequenas variações de escala, as faces a serem classificadas são submetidas a um pré-processamento automático de zoom. Esse pré-processamento aplica, sobre cada face, um zoom de aproximação e de afastamento de 5%, criando uma face a mais a cada operação de zoom. Foi utilizado o valor de 5% porque os testes realizados na etapa de zoom manual mostraram-se promissores para este valor. Assim, para cada face a ser classificada, haverá três imagens: a original, uma ampliada e uma diminuída. No zoom de aproximação ocorre um aumento em ambas as dimensões da imagem. Para manter as dimensões originais da imagem, parte da borda é removida simetricamente. Já no zoom de afastamento não é possível manter as dimensões originais da imagem, trabalhando-se com a imagem em dimensões reduzidas.

A Figura 23 ilustra a aplicação deste pré-processamento sobre uma imagem do banco de dados ORL. A seleção de atributos é realizada da mesma forma, sempre utilizando regiões do mesmo tamanho, independentemente do tamanho da imagem. Já na etapa classificação é realizada uma classificação para cada uma das três faces de entrada, e então é selecionado o resultado que obteve a menor distância. Utilizando essa nova abordagem, realizou-se o teste do deixe-um-de-fora.

(a) Imagem reduzida (b) Imagem original (c) Imagem aumentada

Figura 23: Pessoa 19, na pose 9, e a aplicação do zoom.

Este teste foi aplicado apenas sobre o banco de faces ORL.

4.7.3. Métodos de Avaliação do SDRF

Para validar o sistema proposto de detecção e reconhecimento foram realizados dois testes sobre os vídeos do banco de dados UFPB. O primeiro teste analisou a taxa de reconhecimento sobre os vídeos individuais de cada pessoa do banco. O segundo teste analisou o reconhecimento quando utilizando o vídeo coletivo que contém diversas pessoas do banco em um ambiente de trabalho.

4.7.3.1. Análise dos Vídeos Individuais

Neste teste, é realizada a detecção das faces, utilizando o Módulo Detector, sobre o vídeo de cada uma das pessoas do banco de dados UFPB-Vídeo. Este módulo é configurado para retornar, a partir das faces detectadas, a sub-região três, a qual contém apenas a região de sobrancelhas, olhos, nariz, excluindo a boca. Essa sub-região é entregue ao Módulo Reconhecedor, o qual a redimensiona para ficar compatível com as imagens do banco de faces UFPB-Fotos-3. Em seguida é realizado o reconhecimento, a partir do treinamento previamente realizado com as imagens do banco de dados UFPB-Fotos-3. Contudo, as imagens do banco de dados UFPB-Fotos foram geradas a partir do banco de dados UFPB-Vídeo. Para não utilizar as faces de treinamento na classificação, estas foram retiradas do vídeo.

Esse processo de classificação pode ser considerado tendencioso, pois mesmo removendo as faces utilizadas no treinamento, as condições de iluminação e pose estão muito próximas entre as faces de treino e classificação. Mesmo assim, o teste é válido e permite verificar se os parâmetros de configuração do reconhecedor estão corretos. Caso a taxa de

reconhecimento não esteja acima do limiar pretendido, os parâmetros devem ser re-configurados.

A forma de reconhecimento foi realizada em dois modos. No primeiro, cada quadro do vídeo é analisado isoladamente, e verifica-se se a classificação foi correta ou não. No segundo, o reconhecimento do quadro atual depende do reconhecimento dos *n* quadros anteriores. É então, calculada a moda dos últimos *n* resultados, ou seja, o resultado mais freqüente entre os *n* últimos resultados. A taxa de reconhecimento, para ambos os modos, é então calculada como o total de faces corretamente reconhecidas dividido pelo total de faces detectadas no vídeo. Para obtenção da taxa de reconhecimento global entre todas as pessoas foi realizada a média entre todos os resultados individuais.

4.7.3.2. Análise dos Vídeos Coletivos

Para cada quadro do vídeo foi aplicado o Módulo Detector. Em seguida, sobre as faces detectadas, foi aplicado o Módulo Reconhecedor, previamente treinado com as faces do banco de dados UFPB-Fotos. Neste teste, o vídeo de classificação é totalmente diferente das imagens de faces utilizadas no treinamento, o que torna a tarefa de reconhecimento bem mais difícil do que no teste sobre os vídeos individuais.

O reconhecimento foi realizado quadro a quadro sem levar em consideração os quadros anteriores. A taxa de reconhecimento foi calculada manualmente, com o total de acertos de cada pessoa dividido pelo total de faces detectadas da mesma pessoa. Depois, foi realizada uma média de todos os resultados individuais.

Capítulo 5

Resultados

Este capítulo apresenta os resultados do sistema de uma forma geral, e de forma específica sobre os módulos detector e reconhecedor de faces.

5.1. Detector de Faces

Realizou-se a classificação do detector de faces sobre o banco de dados CMU/MIT [57], e atingiu-se taxa de acerto de 91,2%. Esta taxa é compatível com as encontradas nos detectores mais atuais [33][41][47][48][60][67]. Esse teste foi realizado utilizando como parâmetros o tamanho mínino de faces igual a 20x20 pixels, e a razão de zoom igual a 1,2. Esse valores foram escolhidos baseados nos valores utilizados nos trabalhos de Viola e Jones [67] e de Rowley *et. al* [57].

5.2. Reconhecedor de Faces

Esta seção apresenta os resultados da aplicação dos métodos de avaliação explicados no capítulo anterior, sobre os bancos de dados ORL e UFPB-Fotos.

5.2.1. Validação Cruzada

Os gráficos da Figura 24 e da Figura 25 exibem a performance do sistema durante a aplicação da técnica de validação cruzada deixe-um-de-fora sobre o banco de faces ORL. A Figura 24 exibe os resultados para um número de coeficientes selecionados variando de 1 até 8464. Este último representa o maior valor analisado, visto que as imagens do banco ORL

possuem dimensão de 92x112 e 8464 é a quantidade de coeficientes em um quadrado de dimensões 92x92. Já a Figura 25 destaca a região onde ocorreram os resultados mais significativos, entre 1 e 200 coeficientes.

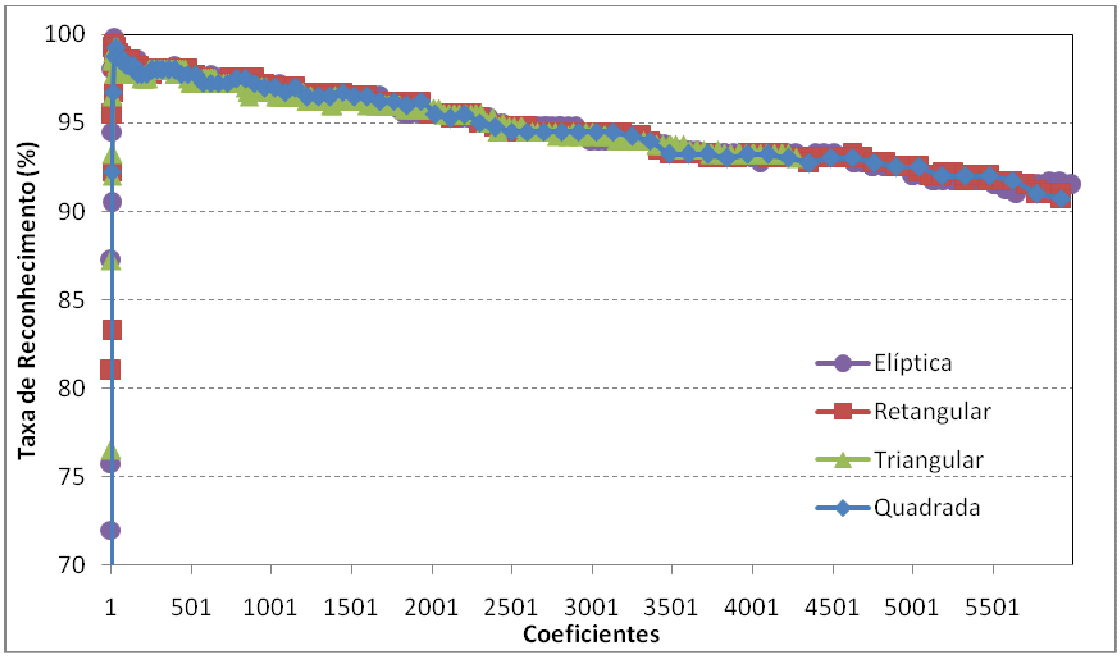


Figura 24: Gráfico da quantidade de coeficientes DCT x taxa do reconhecimento. A taxa varia de 70 a 100 e os coeficientes de 1 a 6000. Banco de faces ORL.

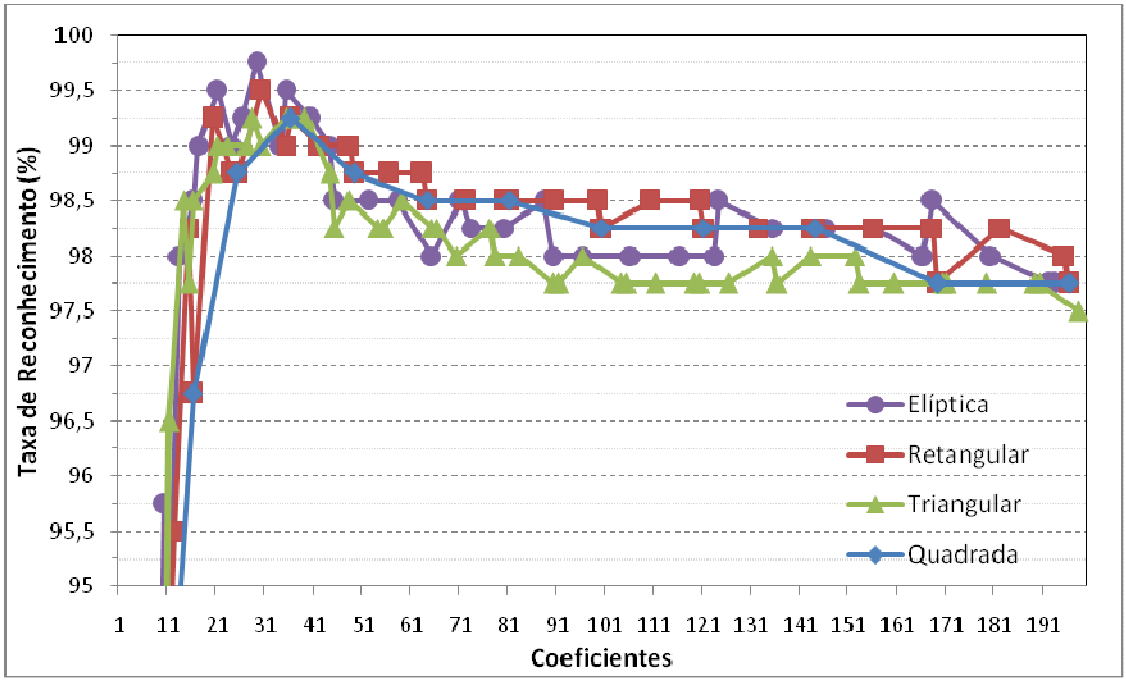


Figura 25: Gráfico da quantidade de coeficientes DCT x taxa do reconhecimento. A taxa varia de 95 a 100 e os coeficientes de 1 a 200. Banco de faces ORL.

Os resultados estão separados pelas quatro formas geométricas utilizadas na seleção de atributos, e é utilizado o classificador do Vizinho Mais Próximo. Pode ser observado que

existe um decaimento na precisão do reconhecimento à medida que ocorre o aumento da quantidade de coeficientes. Isto ocorre devido ao Problema da Dimensionalidade, que diz que o desempenho de um classificador depende da inter-relação entre a quantidade de atributos selecionados (dimensão do espaço) e a quantidade de amostras das classes.

Na Figura 25, pode-se perceber que a forma elíptica possui o melhor resultado: apenas 29 coeficientes DCT são suficientes para atingir uma precisão de 99,75%. Esta taxa representa apenas um erro em 400 classificações no banco de dados ORL. Este resultado foi obtido usando uma forma elíptica com semi-eixos de 6 e 7 pixels.

A região retangular atingiu taxas superiores às regiões quadradas, chegando a 99,50%, enquanto a quadrada só atinge 99,25%. Obviamente, a região quadrada é um caso particular da região retangular, mas os resultados com a região quadrada estão sendo exibidos para a comparação com trabalhos anteriores, como o de Matos *et al.* [46]. Utilizando-se formas elípticas, a taxa de acerto máxima foi de 99,75 %. Fazendo um comparativo ao trabalho de Matos tem-se uma redução da taxa de erro de 0,75% para apenas 0,25%. Esta foi uma redução substancial em 66% da taxa de erro.

Sobre o banco de dados UFPB-Fotos foram realizados testes utilizando apenas a forma geométrica elíptica. Nos testes analisou-se a utilização dos três grupos de fotos do banco de faces UFPB-Fotos. Foram utilizadas imagens de tamanho 128x128 pixels para o UFPB-Fotos-1, 80x108 para o banco de faces UFPB-Fotos-2 e 80x54 para o banco de faces UFPB-Fotos-3. Nestes três bancos de dados as faces estão com o mesmo tamanho, mas o tamanho da imagem varia porque o UFPB-Fotos-2 é uma sub-imagem do UFP-Fotos-1, e o UFPB-Fotos-3 é uma sub-imagem do UFPB-Fotos2. Os melhores resultados foram obtidos com o banco de dados UFPB-Fotos-2, seguidos pelos resultados do banco de dados UFPB-Fotos-1, e por último o banco de dados UFPB-Fotos-3. O gráfico da Figura 26 exibe as taxas de reconhecimento para cada banco *versus* quantidade de coeficientes utilizados. O melhor resultado obtido foi de 98,5% utilizando 73 coeficientes delimitados por uma forma elíptica de eixos com 10 e 10 pixels, ou seja, um círculo. Esse resultado implica 12 erros em 800 classificações. Comparado aos resultados obtidos com o banco de dados ORL, o banco de dados UFPB-Fotos exige uma maior quantidade de coeficientes para atingir o seu melhor resultado. Isto possivelmente acontece porque o banco de faces UFPB-Fotos possui menos background que o banco de faces ORL, e o background influência bastante no nível de cinza médio da imagem. Como o valor do primeiro coeficiente da DCT depende apenas do brilho médio da imagem, então, em imagens com muito background (ORL) o coeficiente DC

(primeiro coeficiente) da DCT tem uma maior importância do que em imagens com pouco background (UFPB-Fotos). Então, como o banco de dados UFPB-Fotos possui menos background que o banco de dados ORL, é necessário que o banco de dados UFPB-Fotos utilize mais coeficientes para compensar uma menor influência do coeficiente DC.

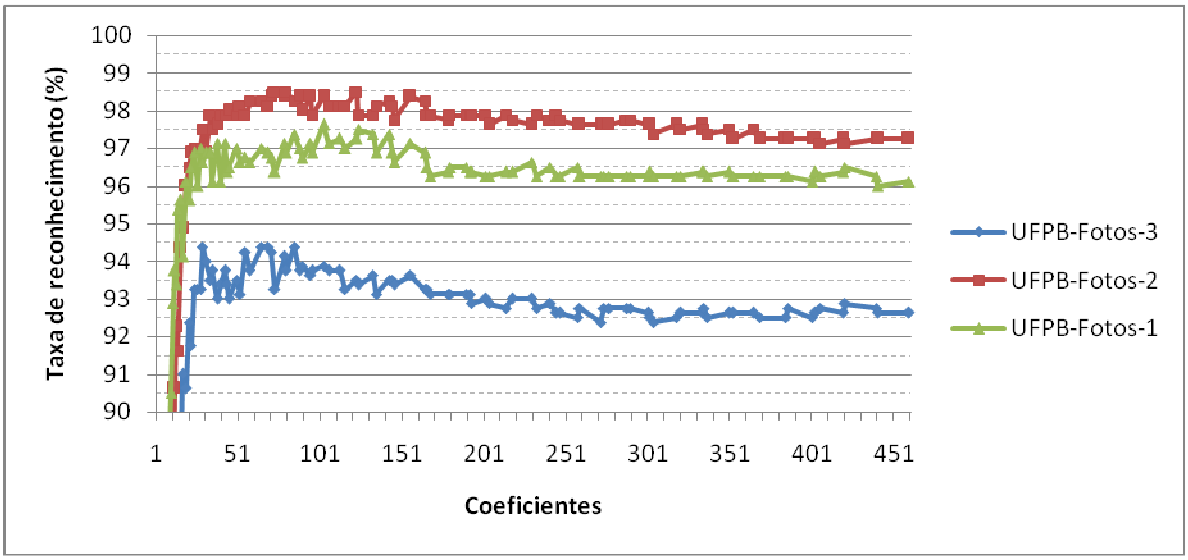


Figura 26: Gráfico da quantidade de coeficientes DCT x taxa do reconhecimento. Taxa variando de 90 a 100 e coeficientes de 1 a 460. Banco de faces UFPB-Fotos.

O banco de faces UFPB-Fotos-2 mostrou resultados superiores ao banco UFPB-Fotos-1. Isto ocorreu porque este último possui muito *background*, o que dificulta o reconhecimento, enquanto o UFPB-Fotos-2 teve parte do background removido. Já o banco de faces UFPB-3 praticamente não possui background, entretanto teve o pior resultado entre os três, pois a região está excessivamente reduzida, o que não permitiu uma diferenciação tão adequada quanto com os outros bancos. Contudo, os três bancos apresentaram resultados superiores a 94%.

5.2.2. Validação Cruzada Acumulativa

Usando a abordagem de validação cruzada acumulativa, o sistema atinge 100% de precisão quando a face correta está dentro de um intervalo de seis faces retornadas. Esta faixa de seis faces foi obtida para as regiões retangulares e triangulares. A região elíptica atingiu 100% apenas no intervalo de sete faces retornadas. Este resultado não foi o esperado, pois as regiões elípticas tinham se mostrado mais eficientes na abordagem deixe-um-de-fora tradicional. Já a região quadrada alcançou 100% com nove faces retornadas. A Figura 27 ilustra estes resultados.

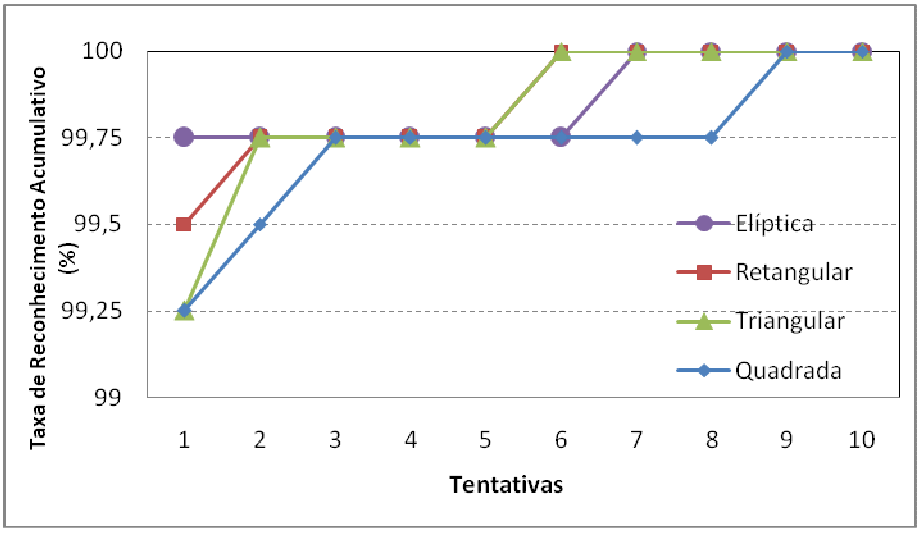


Figura 27: Reconhecimento acumulativo sobre o banco de dados ORL.

É interessante perceber que todas as quatro formas alcançaram um reconhecimento superior a 99% apenas com a primeira face retornada. Os resultados com apenas uma face retornada também podem ser percebidos na Figura 25, selecionando apenas os melhores resultados para cada forma geométrica.

5.2.3. Zoom Automático

A abordagem de ajuste automático de zoom sobre o banco de faces ORL não conseguiu melhorar os resultados da validação cruzada original. O melhor resultado obtido em ambas as abordagens ocorreu utilizando regiões elípticas com semi-eixos de 6 e 7 pixels, contendo 29 coeficientes. Na validação cruzada original atingiu-se o resultado de 99,75% (um erro) e após o ajuste automático de zoom apenas 99,5% (dois erros). O erro que já havia sido detectado na validação cruzada se manteve e um novo erro foi inserido.

O erro comum às duas abordagens ocorreu entre a pessoa 19, em sua pose 9, que foi atribuída à pessoa 11, sendo mais parecida com ela na pose 5. Este erro continuou porque, mesmo aplicando a operação de zoom às imagens a serem classificadas, a distância entre a DCT da imagem original (sem zoom) da pessoa 19, pose 9 e a DCT da imagem da pessoa 11, pose 5 foi menor do que a distância entre a DCT da imagem aumentada (zoom +5%) da pessoa 19, pose 9 para as outras DCTs de imagens da pessoa 19. Desta forma a pessoa 19, pose 9 continuou sendo atribuída à pessoa 11.

5.2.4. Tempo de Processamento

Foi verificado ainda o tempo de execução do módulo de reconhecimento sobre o banco de faces ORL. Duas diferentes implementações foram analisadas: uma desenvolvida em Java e a outra desenvolvida em C++, ambas implementadas neste trabalho.

A maior parte do tempo de execução é gasto com a geração das DCTs das faces do banco de dados - aproximadamente 99% do tempo de execução. No entanto, esta etapa pode ser realizada apenas uma vez e seu resultado armazenado para uma utilização futura. A Tabela 2 mostra o tempo de execução da abordagem de teste deixe-um-de-fora sobre o banco de dados ORL, utilizando a implementação C++, e a Tabela 3 mostra o tempo de execução utilizando a implementação Java. A segunda coluna dessas tabelas mostra o tempo de execução quando as DCTs são completamente geradas, e a terceira coluna mostra o tempo quando as DCTs já estão pré-computadas e são apenas carregadas a partir do disco. A segunda linha das tabelas possui os tempos de execução da aplicação da abordagem deixe-um-de-fora completa, realizando 400 classificações. Já a terceira linha possui os tempos de classificação para apenas uma rodada do deixe-um-de-fora, ou seja, é o tempo da aplicação completa do deixe-um-de-fora dividido por 400.

	Tempo de execução (segundos)	
	Gerando as DCTs	Carregando as DCTs
Obtenção das DCTs	15.660030	0.076801
Abordagem deixe-um-de-fora	0.119613	0.119613
Classificação de uma face	0.000299	0.000299
Tempo total	15.779643	0.196414

Tabela 1: Tempo de execução em C++.

	Tempo de execução (segundos)	
	Gerando as DCTs	Carregando as DCTs
Obtenção das DCTs	42,400000	0,789000
Abordagem deixe-um-de-fora	0,202000	0,202000
Classificação de uma face	0,000505	0,000505
Tempo total	42,602000	0,991000

Tabela 2: Tempo de execução em Java.

Pode-se observar que a versão C++ é mais rápida em todos os aspectos, e é cerca de 10 vezes mais rápida na operação de obtenção das DCT quando em modo de carregamento. Assim, a versão C++ é mais adequada para aplicações em tempo real.

5.3. Sistema SDRF

Esta seção apresenta os resultados da aplicação dos métodos de avaliação, explicados no capítulo anterior, sobre o SDRF, utilizando o banco de faces UFPB-Vídeo.

5.3.1. Vídeos Individuais

Os vídeos individuais foram testados utilizando com base de dados os bancos de faces UFPB-Fotos-1, 2, e 3, entretanto os melhores resultados foram obtidos com o banco de faces UFPB-Fotos-3. Isto ocorreu porque utilizando este banco, o Módulo Detector é configurado para extrair a sub-região 3. Esta sub-região, por ser reduzida e localizada no centro da face, possui menos variações na medida em que a angulação da face varia. Já nos bancos de faces UFPB-Fotos-1 e 2, movimentos de rotação produzem muita mudança no background o que dificulta o reconhecimento. Por isso, foram realizados testes detalhados apenas utilizando o banco de faces UFPB-Fotos-3 e extraindo-se as sub-região três da face a ser classificada.

Os resultados sobre os vídeos individuais do banco de dados UFPB foram promissores. Para este teste foram utilizados os parâmetros de configuração que geraram os melhores resultados nos testes de validação cruzada do banco de dados UFPB-Fotos-3. Esses parâmetros foram: regiões de faces de tamanho 80x54 (obtidas do banco de dados UFPB-Fotos-3), e região de seleção de atributos elíptica de 6x5, contendo 25 coeficientes.

Utilizando essa configuração foram obtidas taxas de acerto 80,17% quando analisando os quadros de vídeo individualmente e de 95,76% quando calculando o resultado através da moda dos últimos 45 resultados. Foram testados outros valores para o tamanho da moda, de 0 até 100 utilizando um passo de 5, entretanto o melhor resultado foi obtido com o valor de 45. O gráfico da Figura 28 exibe o resultado do reconhecimento sobre cada vídeo individual, quando utilizando a moda. Onde cada coluna representa o reconhecimento de uma pessoa, e a última coluna indica a média de todos os resultados. Pode-se perceber que em muitas pessoas o resultado atingiu 100%, e a menor taxa de reconhecimento obtida foi de 72,35%.

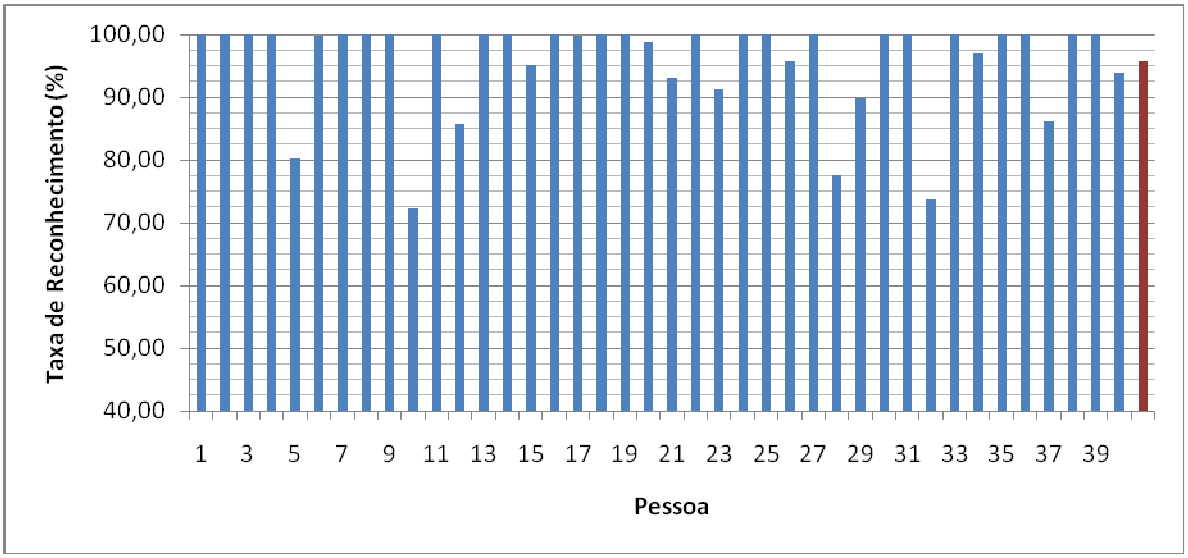


Figura 28: Gráfico da taxa de reconhecimento individual utilizando a moda.

Na Figura 29 tem-se o resultado de quando o reconhecimento é realizado analisando os quadros de vídeo individualmente. Em apenas uma pessoa foi atingido 100% de acerto, e o pior resultado foi de 57,62%. Fica bem visível o melhor desempenho da abordagem que utiliza a moda dos últimos quadros quando comparada ao resultado sem a moda. Pode-se perceber que o resultado, quando utilizando a moda, mostrou um desempenho superior para todas as pessoas. O resultado da pessoa 21, por exemplo, mostrou um ganho de 56,01% para 93,17% com o uso da moda.

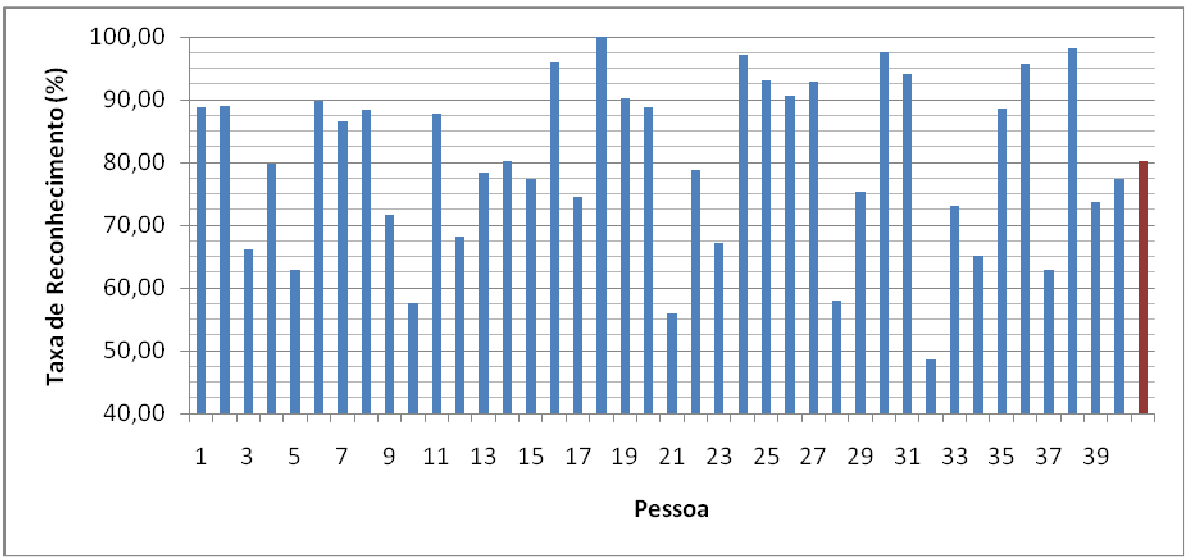


Figura 29: Gráfico da taxa de reconhecimento individual analisando os quadros de vídeo individualmente.

5.3.2. Vídeos Coletivos

Os resultados utilizando o vídeo coletivo do banco de faces UFPB-Vídeos não foram promissores. Apesar de o vídeo ser de tamanho 1920x1080 pixels, a filmagem possui movimentos bruscos da câmera o que gerou algumas imagens borradas, as faces detectadas apresentam um background variado, e as condições de iluminação estão muito diferentes das imagens de treinamento. Algumas faces estão rotacionadas interferindo na extração da sub-região de face. Foi obtida uma taxa de acerto inferior a 5% neste vídeo.

Discussão e Conclusão

Este trabalho propõe um sistema automático de detecção e reconhecimento de face. O desenvolvimento do trabalho atingiu os objetivos propostos, pois foram realizados estudos sobre os métodos presentes na literatura, foi desenvolvido o SDRF, e este foi validado através de testes de desempenho e comparações com outros métodos.

O Módulo Detector apresentou excelentes resultados, detectando faces de forma eficiente e eficaz. Esse bom resultado já era esperado, visto que a técnica utilizada já fora amplamente testada em outros trabalhos e se consolidou como uma das melhores formas de detecção de faces da atualidade.

O Módulo Reconhecedor, no seu melhor resultado sobre o banco de faces ORL, atingiu uma taxa de reconhecimento de 99,75%. Este resultado representa apenas um erro em 400 classificações, e foi obtido utilizando apenas 29 coeficientes da DCT. Este resultado foi superior aos demais trabalhos que utilizam apenas DCT para reconhecer faces. Utilizando abordagem similar de seleção de atributos e classificação, Matos *et al.* [46] obteve uma taxa de reconhecimento de 99,25%, e Hafed e Levine [26] uma taxa de 93,75%. No presente trabalho a taxa de erros foi de 0,25%, enquanto no trabalho de Matos [46], esta taxa foi de 0,75%, houve uma redução substancial de 66,66% na taxa de erro. Testes sobre o banco de faces UFPB-Fotos também foram promissores atingindo taxa de acerto de 98,5%, utilizando 73 coeficientes. Outros métodos de reconhecimento conseguem atingir 100% de acerto sobre o banco de faces ORL, como o método de Kumar et al. [39] e Bicego et al. [9], entretanto, eles utilizam a DCT em conjunto com Modelos Ocultos de Markov, o que leva a um sistema com alto custo computacional.

O novo método proposto de seleção de atributos por baixa freqüência utilizando formas geométricas elípticas mostrou-se superior à seleção utilizando as formas geométricas de retângulos e triângulos. Esse resultado obteve um melhor desempenho, pois a forma elíptica adapta-se melhor à forma como a energia da DCT se expande.

O sistema completo de reconhecimento e detecção desenvolvido neste trabalho mostrou-se eficiente, conseguindo trabalhar em tempo real em vídeos de tamanho 800x600

pixels, detectando e reconhecendo múltiplas faces. Quando aplicado sobre vídeos com apenas uma pessoa do banco de faces UFPB-Vídeo, o sistema atingiu taxa de acerto 80,17% analisando os quadros de vídeo individualmente e de 95,76% quando calculando a moda dos resultados dos quadros anteriores. Já quando aplicado a vídeos coletivos com múltiplas pessoas do banco de faces UFPB-Vídeo, o resultado não foi promissor, porque nestes vídeos os indivíduos se movimentam bastante, os rostos possuem diversas rotações, bem como alterações de iluminação, além da filmagem possuir movimentos bruscos que diminuem a qualidade do vídeo.

Pelos resultados apresentados sobre os bancos de faces de imagens ORL e UFPB-Fotos, pode-se concluir que para reconhecimento de face a aplicação da DCT seguida pela seleção de atributos por baixa freqüência, e utilizando o classificador do vizinho mais próximo é um método apropriado para reconhecimento de face. Entretanto, ele ainda deve ser aperfeiçoado para suprir as necessidades do reconhecimento de face em vídeo com múltiplas pessoas.

Propõe-se como trabalho futuro, a extensão do uso da moda para vídeos contendo diversas pessoas, analisando o resultado de moda individual para cada pessoa. Pretende-se também realizar a investigação da aplicação de um filtro passa-baixas não ideal para minimizar o fenômeno de Gibbs causado pela seleção de atributos do Módulo Reconhecedor.

Referências

[1] ABATE, A. F; NAPPI, M.; RICCIO D.; SABATINO, G. **2D and 3D Face Recognition: A Survey, Pattern Recognition Letter 28**, pp. 1885-1906, 2007.

[2] ABDALLAH, A. S.; ABBOTT, A. L.; EL-NASR, M. A. **A New Face Detection Technique using 2D DCT and Self Organizing Feature Map.** Proceedings of World Academy of Science, Engineering and Technology. vol. 21, 2007.

[3] ANSARI, A.; ABDEL-MOTTALEB, M. **3-D face modeling using two views and a generic face model with application to 3-D face recognition**. In: Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance (AVSS) Miami, FL, USA, pp. 37–44. July 2003.

[4] AT&T Laboratories, Cambridge, UK. **The ORL Database of Faces (now AT&T The Database of Faces)**. Disponível em: <http://www.cl.cam.ac.uk/research/dtg/attarchive/pub/data/att_faces.zip>. Acesso em: 20 jan. 2009.

[5] BAEK, K.; CHANG, Y.; KIM, D.; KIM, Y.; LEE, B.; CHUNG, H.; HAN, Y.; HAHN, H. **Face Region Detection Using DCT and Homomorphic Filter**. In: Proceedings of the 6th WSEAS International Conference on Signal Processing, Robotics and Automation World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, 7-12, Corfu Island, Greece, February 16 - 19, 2007.

[6] BARTLETT, M, S.; MOVELLAN, J. R.; SEJNOWSKI, T. J. **Face Recognition by Independent Component Analysis,** IEEE Transactions on Neural Networks, vol. 13, nº 6, November , 2002.

[7] BATISTA, L. V. **Compressão de Sinais Eletrocardiográficos Baseada na Transformada Cosseno Discreta**. Tese de Doutorado, Pós-Graduação em Engenharia Elétrica, UFPB, Campina Grande, Brasil, 2002.

[8] BELHUMEUR, P. N.; HESOANHA, J. P.; KRIEGMAN, D, J. **Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection**, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 19, no 7, July 1997.

[9] BICEGO, M.; CASTELLANI, U.; MURINO, V. **Using HMM and Wavelets for Face Recognition**. Proceedings of the 12th International Conference on Image Analysis and Processing, IEEE, 2003.

[10] BHUIYANV, M. A. A.; AMPORNARAMVETH, V.; MUTO, S. Y.; UENO, H. **Face Detection and Facial Feature Localization for Human-machine Interface**. NII Journal No.5, 2003.

[11] BRADSKI, G. R.; KAEHLER, A. **Learning OpenCV: Computer Vision with the OpenCV Library**. O'Reilly, 2008.

[12] CAMPOS, T. E, **Técnicas de Seleção de Características com Aplicação em Reconhecimento de Faces**. Dissertação de Mestrado, USP, São Paulo, Brasil, 2001.

[13] CHAI, D. and WONG, K. W. **Facial Image Processing: An Overview,** Proceeding of the IEEE Conference on Cybernetics and Intelligent Systems, Singapore, 2004.

[14] CHIEN, J. T.; WU, C. C. **Discriminant Waveletfaces and Nearest Feature Classifiers for Face Recognition**. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, nº 12, pp. 1644-1649, Dec. 2002. DOI: 10.1109/TPAMI.2002.1114855 .

[15] CHELLAPPA, R; WILSON, C. L.; SIROHEY, S. **Human and Machine Recognition of Faces: A Survey**. Proceedings of IEEE, vol. 83, nº 5 (703-740), May, 1995.

[16] DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**. Second Edition, Wiley-Interscience, 2000.

[17] EKENEL, H. K.; GOA, S. H.; FISCHERM, M.; STIEFELHAGEN, R. **Face Recognition for Smart Interactions,** IEEE ICME, 2007.

[18] FISHER, R. A. **The Statistical Utilization of Multiple Measurement**. In: Annals of Eugenics, pp. 376-386, 1938.

[19] FLEURET, F.; GEMAN, D. **Fast Face Detection with Precise Pose Estimation**. Pattern Recognition, 2002. Proceedings. 16th International Conference on, vol.1, no. 1, pp. 235-238, 2002.

[20] FREUND, Y.; SCHAPIRE, R. E. **A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting**. Journal of Computer and System Sciences, Volume 55, Issue 1, Pages 119-139, August 1997.

[21] FROBA, B. and ERNST, A. **Face Detection with the Modified Census Transform**. Automatic Face and Gesture Recognition. Proceedings. Sixth IEEE International Conference, pp. 91-96, 17-19 May 2004.

[22] GARCIA, C. and DELAKIS, M. **Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection.** IEEE Trans. Pattern Anal. Mach. Intell. 26, pp. 1408-1423, 11 Nov. 2004. DOI: 10.1109/TPAMI.2004.97.

[23] GIBBS, J. W., **Fourier Series. Nature 59**, 200 (1898) and 606 (1899).

[24] GONZALEZ, R. C. and WOODS, R. R. **Digital Image Processing Using Matlab**. 3ed. Prentice Hall, 2007.

[25] HAYKIN, Simon. **Neural Networks And Learning Machines**. 3rd Edition. Prentice Hall, 2008.

[26] HAFED, Z. M and LEVINE, M. D. **Face Recognition Using Discrete Cosine Transform**. International Journal of Computer Vision, v. 43(3), pp. 167-188, 2001.

[27] HEISELE, B.; POGGIO, T.; PONTIL, M. **Face Detection in Still Gray Images**. AI Memo 1687, Center for Biological and Computational Learning, MIT, Cambridge, MA, 2000.

[28] HONÓRIO, T. C. S.; DUARTE, R. C. M.; NOBRE NETO, F. D.; ALMEIDA, T. P.; BATISTA, L. V. **Classificação de Texturas Usando PPM**. In: Reunião Anual da SBPC, Belém. Anais/Resumos da 59a Reunião Anual da SBPC: publicação eletrônica, 2007.

[29] ISO/IEC 13818-1. Information technology. **Generic coding of moving pictures and associated audio information**: Part 1: Systems, 2000.

[30] IVANCEVIC, V., KAINE, A. K.; MCLINDIN, B. SUNDE, J. **Factor Analysis of Essential Facial Features**. Proceedings of the 25th International Conference on Information Technology Interfaces, pp. 187-191, Zegreb, Croatia, (D Simic, Ed), 2003.

[31] JAIN, A. K.; DUIN, R. P.W.; MAO, J. **Statistical Pattern Recognition: A Review**. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 4-37, January, 2000.

[32] JAIN, A.; HONG, L.; PANKANTI, S. **Biometric Identification**. Commun. ACM 43, 90-98. 2 Feb. 2000. DOI: 10.1145/328236.328110.

[33] JIANXIN W.; BRUBAKER, S.C.; MULLIN, M.D.; REHG, J.M. **Fast Asymmetric Learning for Cascade Face Detection**. Pattern Analysis and Machine Intelligence, IEEE Transactions on vol.30, no. 3, pp. 369-382, March 2008.

[34] JONES, M. J. and VIOLA, P. **Face Recognition Using Boosted Local Features**. IEEE International Conference on Computer Vision, 2003.

[35] KANADE, T. **Computer Recognition Of Human Faces**. Birkhauser, Basel, Switzerland, and Stuttgart, Germany, 1973.

[36] KELLY, M. D. **Visual Identification of People by Computer**. Tech. rep. AI-130, Stanford AI Project, Stanford, CA, 1970.

[37] KERVRANN, C.; DAVOINE, F.; PÉREZ, P.; FORCHHEIMER, R.; LABIT, C. **Generalized Likelihood Ratio-Based Face Detection and Extraction of Mouth Features**, Pattern Recog. Lett. 18, 1997.

[38] KHAN, A.S., ALIZAI, L.K. **Introduction to Face Detection Using Eigenfaces**, Emerging Technologies. ICET apos. pp: 128-132, 2006. DOI: 10.1109/ICET.2006.335908.

[39] KUMAR, S. A. S.; DEEPTI, D. R.; PRABHAKAR, B. **Face Recognition Using Pseudo-2D Ergodic HMM**, IEEE, ICASSP 2006.

[40] KOVAC, J.; PEER, P.; SOLINA, F. **Human Skin Color Clustering For Face Detection.** EUROCON 2003. Computer as a Tool. The IEEE Region 8, vol.2, pp. 144-148 vol.2, 22-24, Sept. 2003.

[41] LI, S.Z.; ZHENQIU Z. **FloatBoost learning and statistical face detection**, Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.26, no.9, pp.1112-1123, September 2004.

[42] LIN, S. H.; KUNG, S. Y.; LIN, L. J. **Face recognition/detection by probabilistic decision-based neural network**, Neural Networks, IEEE Transactions on , vol.8, no.1, pp.114-132, Jan 1997.

[43] MACQUEEN, J. B. **Some Methods for classification and Analysis of Multivariate Observations**. In Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. 1: 281–297, University of California Press, 1967.

[44] MARQUES, J. R. T.; POEL, J. V. D.; BATISTA, L. V.; GUILHERME C. G. **Mammographic Image Noisy Bit Planes Identification and Removal Using a Binary PPM Algorithm and an Open Source Architecture**. In: XXI Congresso Brasileiro de Engenharia Biomédica, Salvador - BA. Anais do XXI Congresso Brasileiro de Engenharia Biomédica - CBEB 2008.

[45] MATOS, F. M. S. **Reconhecimento de Faces Utilizando Seleção de Coeficientes da Transformada Cosseno Discreta**. Dissertação de Mestrado - Curso de Pós-Graduação em Informática, UFPB, João Pessoa, 2008.

[46] MATOS, F. M. S.; BATISTA, L. V.; POEL, J. V. D. **Face Recognition Using DCT Coefficients Selection**, Proceedings of the 23rd Annual ACM Symposium on Applied Computing. Fortaleza, Brazil. 16-20 March 2008.

[47] MEYNET, J.; POPOVICI, V.; THIRAN, J. P. **Face detection with boosted Gaussian features**, Pattern Recognition, Volume 40, Issue 8, Part Special Issue on Visual Information Processing, pp. 2283-2291, August 2007. DOI: 10.1016/j.patcog.2007.02.001.

[48] MITA, T.; KANEKO, T.; HORI, O. **Joint Haar-like Features for Face Detection**. In Proceedings of the Tenth IEEE international Conference on Computer Vision - Volume 2 ICCV. IEEE Computer Society, Washington, DC, 1619-1626, October 17 - 20, 2005. DOI: 10.1109/ICCV.2005.129.

[49] NEFIAN, A. V. and HAYES, M.H. **Maximum likelihood training of the embedded HMM for face detection and recognition**. In: International Conference on, vol.1, no., pp.33-36, 2000.

[50] OMAIA, D.; POEL, J.; BATISTA, L. V. **2D-DCT Distance Based Face Recognition Using a Reduced Number of Coefficients.** XXIIth Brazilian Symposium on Computer Graphics and Image Processing, SIBGRAPI, 2009.

[51] OPENCV, **Open Source Computer Vision Library**, Intel. Disponível em: <http://sourceforge.net/projects/opencvlibrary/>. Acesso em: 01 fev. 2009.

[52] PAPAGEORGIOU, C. P.; OREN, M.; POGGIO, T. **A general framework for object detection**. Computer Vision, Sixth International Conference on, vol., no., pp.555-562, 4-7, Jan 1998.

[53] PODILCHUK, C. and ZHANG X. **Face Recognition Using DCT-Based Feature Vectors**, IEEE, 1996.

[54] POPOVICI, V. and THIRAN, J. P. **Face Detection Using an SVM Trained in Eigenfaces Space**. Audio-and Video-Based Biometric Person Authentication 2003. DOI: 10.1007/3-540-44887-X_23.

[55] RAO, K. R. **Discrete Cosine Transform – Algorithms, Advantages, Applications, Academic Press**, Inc, 1990.

[56] RIHAN, J.; KOHLI, P.; TORR; P.H.S. **OBJCUT for Face Detection**. In Proceedings of ICVGIP, pp. 576-584. 2006.

[57] ROWLEY, H A.; BALUJA, S.; KANADE, T. **Neural Network-Based Face Detection**, IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, pp. 23-38. 1998.

[58] RUIZ-DEL-SOLAR, J. and NAVARRETE, P. **Eingenspace-Based face Recogniton**: **A Comparative Stydy of Different Appoaches**, IEEE Transaction on Systems, MAN and Cybernetics – Part C: Applications and Reviews, vol. 35, no 3, August 2005.

[59] SABER, E. and TEKALP, A. M. **Frontal-view face detection and facial feature extraction using color, shape, and symmetry based cost functions.** Pattern Recogn. Lett. 19, 669-680. 8 Jun. 1998. DOI: 10.1016/S0167-8655(98)00044-0.

[60] SAUQUET, T.; RODRIGUEZ, Y.; MARCEL, S. **Multiview Face Detection**, IDIAP, IDIAP-RR, no:49, 2005.

[61] SALOMON, D. **Data Compression the Complete Reference**. Third Edition. Springer, 2004.

[62] SIROVICH, L. and KIRBY, M. **Low-dimensional procedure for the characterization of human faces**. Journal of the Optical Society of America A, vol 4, No.3, pp. 519-524, March 1987.

[63] SUNG, K-K., and POGGIO, T. **Example-based learning for view-based human face detection**, IEEE Trans. Pattern Anal. Mach. Intelligence 20, , pp.39–51, 1998.

[64] TERRILLON, J.C. and AKAMATSU, S. **Comparative performance of different chrominance spaces for color segmentation and detection of human faces in complex scene images**. In: Proceedings of the Vision Interface, pp. 180-197, 1999.

[65] TURK, M.A. and PENTLAND, A.P. **Face recognition using eigenfaces**. Computer Vision and Pattern Recognition, 1991. Proceedings CVPR 91, IEEE Computer Society Conference pp.586-591, 3-6, Jun 1991.

[66] VEZHNEVETS, V.; SAZONOV, V.; ANDREEVA, A. **A survey on pixel-based skin color detection techniques** in Proc. Graphicon, 2003.

[67] VIOLA, Paul; JONES, Michael. **Robust Real-time Face Detection.** International Journal of Computer Vision, 57(2):137-154, 2004.

[68] WANG, J. and SUNG, E. **Frontal-view face detection and facial features extraction using color and morphological operations.** Pattern Recogn. Lett. 20, pp. 1053-1068, 10 Oct. 1999. DOI: 10.1016/S0167-8655(99)00072-0.

[69] WEBB, A. R. **Statistical Pattern Recognition**, Second Edition, John Wiley and Sons Ltd, 2002.

[70] XIAO, R.; LI, M. J.; ZHANG, H. J. **Robust multipose face detection in images.** Circuits and Systems for Video Technology, IEEE Transactions on , vol.14, no.1, pp. 31-41, Jan. 2004.

[71] YANG, M.; KRIEGMAN, D.; AHUJA, N. **Face detection using multimodal density models.** Comput. Vis. Image Underst. 84, 2, 264-284, Nov. 2001. DOI: 10.1006/cviu.2001.093.

[72] ZHAO, W.; CHELLAPPA, R.; PHILLIPS, P. J., ROSENFELD, A. **Face Recognition**: **A Literature Survey**, ACM Computing Surveys, vl. 35, nº 4, pp. 399-458, 2003.

Apêndice I – Artigo Publicado

Como um dos resultados preliminares desse trabalho, o artigo **“2D-DCT Distance Based Face Recognition Using a Reduced Number of Coefficients”** foi publicado nos Anais do XXII Simpósio Brasileiro em Computação Gráfica e Processamento de Imagens (XXII Sibgrapi), a ser realizado no período de 11 a 14 de Outubro de 2009 na cidade do Rio de Janeiro – RJ. A versão final do artigo é apresentada nesse apêndice.

2D-DCT Distance Based Face Recognition Using a Reduced Number of Coefficients

Derzu Omaia
PPGI / DI / UFPB
Brasil
derzu@lavid.ufpb.br

JanKees v. d. Poel
PPGEM / DEM / UFPB
Brasil
jkvdpoel@yahoo.com.br

Leonardo V. Batista
PPGI / DI / UFPB
PPGEM / DEM / UFPB
Brasil
leonardo@di.ufpb.br

ABSTRACT

Automatic face recognition is a challenging problem, since human faces have a complex pattern. This paper presents a technique for recognition of frontal human faces on gray scale images. In this technique, the distance between the Discrete Cosine Transform (DCT) of the face under evaluation and all the DCTs of the faces database are computed. The faces with the shortest distances probably belong to the same person; therefore this evaluating face is attributed to this person. The distance is calculated as the sum of the differences between the modules of DCT coefficients. Only a few coefficients are used in this computation; they are selected from the low frequency of the DCT. Experimental tests on the ORL database reaches a recognition rate of 99.75%, with low computational cost and no preprocessing step. Additionally, the method achieved 100.0% of recognition accuracy when applying a zooming normalization over the ORL database.

KEYWORDS

Face Recognition, Discrete Cosine Transform, Feature Selection, Classification, Performance.

I. INTRODUCTION

Systems for biometric pattern recognition are widely used in security area. These systems use unique human characteristics such as fingerprints, iris, voice, face, enabling differentiation among human beings [14].

Facial research in computer vision can be divided into several areas, such as face recognition, face detection, facial expressions analysis, among others [14]. Face recognition systems have a wide range of application, especially when dealing with security applications, like computer and physical access control, real-time subject identification and authentication, and criminal screening and surveillance. It is also a research topic in several fields, like image processing, neural networks, computational vision, computer graphics and psychology[14].

The biggest difficulty in developing a robust face recognizer is that a human face can undergo several transformations: the same person may use different face accessories like glasses, earrings, piercings and makeup; he can change his hair into long/short/skin/fringe and even dye it; can wear beard and mustache; can be at many facial expressions like smiling, sad or doing a grimace; and he can be at many different ages. These facial changes make the recognition task very difficult – even human beings do some recognition mistakes from time to time.

This article presents a holistic face recognition method which is based on the magnitude of the Discrete Cosine Transform (DCT) coefficients. The most relevant coefficients for recognition are selected using the technique of low frequencies selection [11]. To classify a face, its DCT coefficients are extracted. Next, the distances between the DCT coefficients of the face to be classified and the DCT coefficients of all faces in the database are calculated. The shortest of all distances will probably be associated with faces of the same person; therefore the face under classification will be classified as belonging to the person whose face in the database has the shortest distance to the face to be classified.

The Order-One Minkowski Metric was used to calculate the distances between the face coefficients. This distance is calculated as the sum of the module of the differences between the absolute values of the DCT coefficients and it turns to be simple and efficient.

Several methods described in literature are also based on mathematical transforms, such as DCT, Karhunen-Loève Transform (KLT) also known as Principal Components Analysis (PCA) and the Wavelet Transform [11]. Many DCT-based face recognition methods are being proposed, mainly because of some DCT properties and the existence of fast algorithms to perform its computation. Those methods are achieving high rates of hits, comparable with those obtained by methods based on KLT and Wavelet [6].

Some of the DCT-based methods are not holistic, as they only use blocks of DCTs on specific face positions, such as eyes, nose and mouth [3][10][12]. Others are holistic and use DCT to select the attributes and then use classifiers based on similarities such as the K-Nearest Neighbor and the Minimum Distance to Prototype [9][11]. Hafed and Levine [9] reached a recognition rate of 93.0%. Podilchuk and Xiaoyu [12] obtained a 94% rate of success, whilst Matos [11] reached 99.25%. Despite Kumar et al. [10] and Bicego et al. [3] reached 100%, they use the DCT

in conjunction with Hidden Markov Models (HMMs), which leads to a system with a high computational cost.

The method proposed in this paper uses the DCT not only to generate but also to select the most important attributes of the faces under analysis, and applies the Nearest Neighbor Classifier to classify the test face among all training faces. The work presented here has results that exceeds those exposed in [9] and in [11] and improve the previously obtained results through some changes in the methods.

Using the Olivetti Research Lab (ORL) database and the leave-one-out training/classification approach, a success rate of 99.75% was achieved. This rate represents 1 error in 400 classifications. These results were obtained at a low computational cost, since DCT computing techniques are very effective and the Minkowski Order-One distance is very fast to calculate.

II. DISCRETE COSINE TRANSFORM

The mathematical theory of linear transforms plays a very important role in the signal and image processing area. They generate a set of coefficients from which it is possible to restore the original samples of the signal.

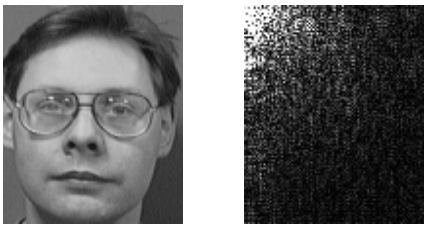
In many situations, a mathematical operation – generally known as a transform – is applied to a signal that is being processed, converting it to the frequency domain. With the signal in the frequency domain, it is processed and, finally, converted back to the original domain. A mathematical transform has an important property: when applied to a signal, i.e., they have the ability to generate decorrelated coefficients, concentrating most of the signal's energy in a reduced number of coefficients [2].

The Discrete Cosine Transform (DCT) is an invertible linear transform that can express a finite sequence of data points in terms of a sum of cosine functions oscillating at different frequencies. The original signal is converted to the frequency domain by applying the direct DCT transform and it is possible to convert back the transformed signal to the original domain by applying the inverse DCT transform.

After the original signal has been transformed, its DCT coefficients reflect the importance of the frequencies that are present in it. The very first coefficient refers to the signal's lowest frequency, known as the DC-coefficient, and usually carries the majority of the relevant (the most representative) information from the original signal. The last coefficient refers to the signal's higher frequencies. These higher frequencies generally represent more detailed or fine information of signal and probably have been caused by noise [7]. The rest of the coefficients (those between the first and the last coefficients) carry different information levels of the original signal.

In the image processing field, it is interesting to use a two-dimensional DCT (2D-DCT), because images are intrinsically two-dimensional elements. The standard JPEG, for example, establishes the use a 2D-DCT at the decorrelation step [13].

Figure 1 shows the application of the DCT on one of the face images obtained from the ORL database. Figure 1.a displays the original image, and Figure 1.b displays the result of applying the DCT on the original image. At Figure 1.b, it is possible to verify that most of the image's energy is concentrated in the upper left corner. This is the region that represents the DCT lowest frequency coefficients.



(a) Original image. (b) Normalized DCT image of the person 4, pose 9.

Figure 1: ORL Database: person 4, pose 1, and it's DCT.

Figure 2 shows the original face and two reconstructed versions of it after applying the DCT. Figure 2.a corresponds to the original face, with dimensions 92x112, i.e., a matrix with 10.304 values. The next two faces represent the reconstruction of the original image using, respectively, 2.576 e 625 DCT coefficients. To obtain the reconstructed images, the following procedure was adopted: application of the direct DCT transform on the original face, then setting to zero the DCT coefficients to be discarded and, finally, application of the inverse DCT transform on the new matrix of coefficients. Figure 2.b illustrates the reconstruction of the original face considering only the DCT coefficients from the first quadrant, that is, preserving only 25% of the DCT coefficients (coefficients from position [1,1] to [56,46] were retained and the rest of them were put to zero). Figure 2.c illustrates the reconstruction of the original face preserving only 6,07% of the DCT coefficients (coefficients from position [1,1] to [25,25] were retained and the rest of them were put to zero).

(a) Original image. (b) Reconstructed image with 25% of the coefficients preserved. (c) Reconstructed image with 6,07% of the coefficients preserved.

Figure 2: ORL Database: Person 4, pose 1, and its image reconstruction.

From what was shown due to the reconstructed images from Figure 2, it is possible to see that the DCT-based dimensionality reduction is capable of producing good results. The reconstructed images when considering only the low frequency coefficients obtained from applying the DCT do present a reduction in the details, but important information to characterize this images (such as the forehead line, the nose, the mouth, the ears, among others) are preserved. These results do even suggest the complete viability of a face recognition method that uses a DCT-based dimensionality reduction.

The 2D-DCT used in this work is the DCT-II. The DCT-II definition is shown in Equations (1) and (2). In this context, the original image is the gray-scale matrix x[m,n], with dimensions m by n, that represents the image. The DCT-II computation then produces a matrix X[k,l], also with dimensions m by n, of coefficients. The variables m and n are the coordinates in the space domain and k and l are the coordinates in the frequency domain [13].

$$X[k,l] = \frac{2}{N} c_k c_l \sum_{m=0}^{a-1} \sum_{n=0}^{b-1} x[m,n] \cos\left[\frac{(2m+1)k\pi}{2N}\right] \cos\left[\frac{(2n+1)l\pi}{2N}\right] \quad \textbf{(1)}$$

Where, in Equation (1):

$$c_k, c_l = \begin{cases} \left(\frac{1}{2}\right)^{1/2} \text{ to } k = 0, l = 0 \\ 1 \text{ to } k = 1,2,\ldots a-1 \text{ and } l = 1,2,\ldots b-1 \end{cases} \quad \textbf{(2)}$$

The first coefficient, X[0,0], is referred as the DC (Direct Current) coefficient and depends only on the average brightness of the image. The other coefficients are known as AC (Alternate Current) factors [13].

In this paper, the DCT used is always the 2D-DCT so, from now on, the term DCT actually means the 2D-DCT.

III. PROPOSED FACE RECOGNITION METHOD

The proposed method is divided into three stages: attribute selection, training and classification. Figure 3 shows a generic face recognition system, where these stages are depicted. The feature selection stage extracts the attributes. The training step chooses the best group of features to perform the identification task, considering all the training samples. The classification stage compares an unknown sample with the ones in the database, using the selected features to determine the best match. In our system no preprocessing stage is performed.

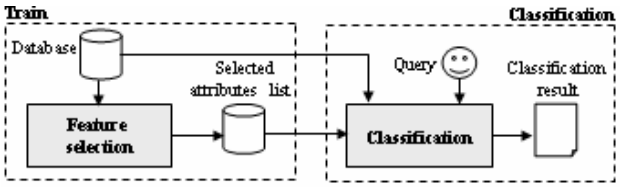
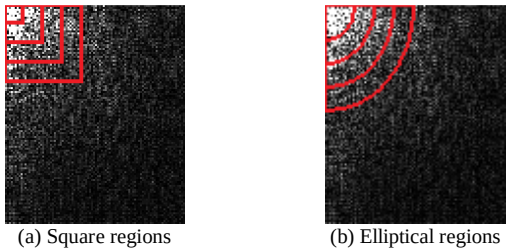


Figure 3: Generic face classifier, adapted from Matos [11].

A. Attribute Selection Stage

Whereas the faces have been previously detected by some other algorithm, the stage of attribute selection specify the list of characteristics that best represents a person, considering the various poses present in the training database. The process of attribute selection extracts the most relevant information to be used to discriminating a person, i.e. the information that is common to a person and is different for the other persons [5].

To perform the attribute extraction, the approach used is the selection of the lowest DCT frequency coefficients. This approach is fast and simple, as it neither evaluates the DCT coefficient of the image, nor performs calculations or any kind of comparison. As the DCT of an image is a 2D signal, the selection of the lower frequency coefficients consists simply in defining a geometric region at the beginning of the 2D signal [11]. An example of such masks for attribute selection using four square regions is shown in Figure 4.a, illustrating this selection approach for low frequencies. It shows the result of applying the 2D-DCT on an image, followed by a normalization stage for easy viewing. As can be seen, the largest energy concentration occurs in the low frequency coefficients region. Thus, this approach is suitable to capture the most important coefficients of the image [11].



(a) Square regions (b) Elliptical regions

Figure 4: Normalized DCT from person 4, pose 1 of the ORL database, and the regions of low frequency.

One of the major contributions of this work is basically on the way to select the most important of the DCT coefficients. This is done by defining a new geometric shape that is to be used to select the attributes. A previous work [11] used square regions on the coefficient selection stage, as in Figure 4.a. But, in the current study, regions with square, rectangular, triangular, and elliptical shapes were tested.

In Figure 4.b we can see that an elliptical shape selects the coefficients of greater amplitude (white dots of the image) better than a square shape, thus selecting the most relevant coefficients of the DCT 2D signal.

B. Training Stage

The training stage performs various attribute selection steps and decides which one produces the best result. With this in mind, the training algorithm tests various geometric

regions in various scales. Tests were performed on square, rectangular, triangular and elliptical regions.

In this paper, the ORL Database of Faces [1] was used to train and test the proposed method. This database contains photos from faces of 40 people, each person in 10 different poses, totalizing 400 photos. The faces were photographed at different moments, with varying lighting, facial expressions (eyes closed/opened, smiling/not smiling), facial poses and facial details (with/without glasses, with/without beard), among other types of variations. The images are in grayscale, with 256 different levels and a dimension of 92x112 pixels. Some examples are shown in Figure 5.

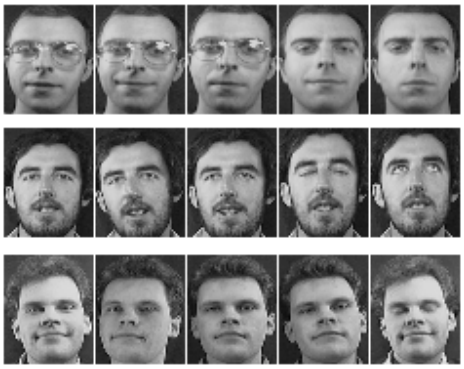


Figure 5: Examples from the ORL database: persons 19, 11 and 36 at their pose 1 to 5.

As the ORL database images are relatively small, tests were performed throughout the first quadrant region of the DCT. This quadrant is the DCT's lower frequency region, and goes from the point (1,1) to (46,56). Over this reduced region of the DCT, a brute force test was used to determine which geometric region and size leads to the best result, which were obtained with elliptical regions of size 6x7 containing 29 coefficients.

These results were obtained over the ORL database. For each different face database, all the training must be repeated, since the choice of the most important group of coefficients is global and not a group for each person. This indicates that, at each insertion of a new person into the database, all the training should be repeated. That happens because the performance of a classifier depends on the interrelationship between the number of selected attributes (space dimension) and the number of samples of class, this problem is known as the Curse of Dimensionality [4].

C. Classification Stage

The classification stage compares a test face with a series of faces from a database, considering the list of attributes selected, and indicates to whom the test face best matches.

Here, the comparison is done using a similarity measure that indicates the resemblance between the test face and all the other faces. The proposed method used, as its similarity measure, the Order-One Minkowski distance between the amplitude of the coefficients of the training poses and the test face coefficients amplitudes. This distance is calculated as the sum of the differences between the modules of the DCT coefficients.

The classification approach used is the Nearest Neighbor (NN). This approach was chosen among others, because it showed the best results in previous works such as Matos [11]. The Nearest Neighbor classifier calculates the distance of the face under evaluation for all other faces (neighbors) in database. The face in question will be attributed to the person belonging to the face with the smallest distance (nearest neighbor). Theoretically, NN can be defined as follows.

Given the m selected coefficients for the person *j*, $\{y_{1j}, y_{2j}, ..., y_{mj}\}$, and the amplitudes of the training coefficients from the person *j* in pose *k*, $\{w_{1jk}, w_{2jk}, ..., w_{mjk}\}$, where w_{ijk} being the coefficient of same position as y_{ij}.

Consider now a test face *f* and its DCT coefficients' amplitudes $\{v_{1f}, v_{2f}, ..., v_{mf}\}$, with v_{if} a coefficient in the same position as y_{ij}.

The distance between a test face *f* and a person *j* in pose k, with $j = 1, 2, \ldots, p$ and $k = 1, 2, \ldots, q$, is given by:

$$D_{NN\,fjk} = \sum_{i=1}^{m} \left| w_{ijk} - v_{if} \right| \quad \textbf{(3)}$$

So, according to Equation (4), a test face *f* is classified as the person *j* when:

$$D_{NN\,fjk} \leq D_{NN\,fgh}, \forall j \neq g, \forall k \neq h \quad \textbf{(4)}$$

The minimum degree of similarity is zero, and that value means that the face being classified is exactly the same of some of the faces in the database. A negative value does not happen because the used distance has absolute values.

IV. TESTS

Tests were conducted to validate the proposed method. The first one was the common leave-one-out approach and the second was the cumulative leave-one-out. Next, a manual zoom test was performed to the single error found in the best result obtained with the method. Then, the procedure involved in this zoom test was automated.

A. Leave-one-out

Cross validation is a technique that analyses how the results of a statistical analysis will generalize to an independent data set. The leave-one-out test rule is a kind of Cross-validation, which leaves one pose out of the training set and do the classification test with it [8]. As the ORL Face Database [1], (which was) used in the tests, has ten poses for each person stored in it, then ten rounds of tests were done. In the first round, the first pose from all

poses is excluded from the training set and used as the test face. In the second round, the second pose is excluded from the training set and used as the test face and so on. This test was performed several times, with different numbers of selected coefficients, selected using different geometric shapes.

B. Cumulative Leave-one-out

The Cumulative Leave-one-out test performed measure the Accumulative Recognition. In this test, the classifier returns a ranked list, in order of increasing distance, of the faces which are most similar to the test face. A common NN classifier would use only the first element of the list (the one with the smaller distance) to perform the classification, The cumulative classifier indicates the minimum size of the list so that the classifier achieves 100%. That is, the minimum size of the list containing a pose of the same person of the test face. This test was also done to different geometric shapes of attribute selection.

C. Manual Zoom

An analysis was performed on the only error found in our best result. In this error, the person 19 in his pose 9 was assigned to person 11, being more like her in pose 5. Figure 6 shows the 10 poses of the person 19 in ascending order. Pose 9 is emphasized and the last image of the second line is the pose 5 of person 11. It can be seen that the pose 9 of person 19 has a slightly smaller face than the nine other poses of this person. Also, it can be seen, that the face of the person 11 in pose 5 has a closest size to person 19 in pose 9 than the rest of the poses of the person 19.

Figure 6: ORL Database: Person 19 in yours 10 poses and person 11 in pose 5.

A test was performed to verify if the recognizer is sensible to the zoom operation. This test manually applied a 5% zoom to the pose 9 of person 19. With this zoom, this pose now has a closer size when compared to the other poses of the person 19, as can be seen in Figure 7. With this manual change, the recognizer reached a hit rate of 100%. That implies that the recognizer is sensible to the zoom operation.

Figure 7: Person 19, in yours 10 poses. The pose 9 has a 5% zoom in.

D. Automatic zooming normalization

Aiming for a face database more robust to small scale changes, the database used was submitted to an automatic pre-processing zoom normalization. This pre-processing applies, on each face, a zoom in and out of 5%, creating one more face for each zoom operation, tripling the size of the database. Thus, each pose will be composed of three images: the original one, the augmented one and the diminished one. Figure 8 illustrates the application of this preprocessing over a sample image of the ORL database. Even with images of slightly different sizes, the attribute selection is performed in the same way, always using regions of the same size regardless of the size of the image.

Using this new stored database, the test using the common leave-one-out procedure is then executed.

(a) Reduced image (b) Original image (c) Augmented image

Figure 8: Person 19, on pose 9, and it's zoom application.

V. RESULTS

This topic presents the results of applying the tests described in the previous section on the ORL database and on the zoomed version of this database.

A. Leave-one-out

Figure 9 shows the system performance over the leave-one-out test, and how the number of DCT coefficients used might have an effect on the overall performance of the recognition system when applied to the ORL database. The recognition rate is calculated as the number of correct hits divided by the number of faces present in the database, the result of this division is then multiplied by 100. The results are separated by the four geometric shapes used on the attributes selection, and was used the Nearest Neighbor classifier. It can be observed that there is a decrease in the recognition accuracy rate as the numbers of coefficients increase. That happens because of the Curse of Dimensionality, which states that the performance of a classifier depends on the interrelationship between the

number of selected attributes and the number of samples of classes, and the excessive increase of the number of coefficients causes efficiency decrease [4]. This recognition rate is calculate

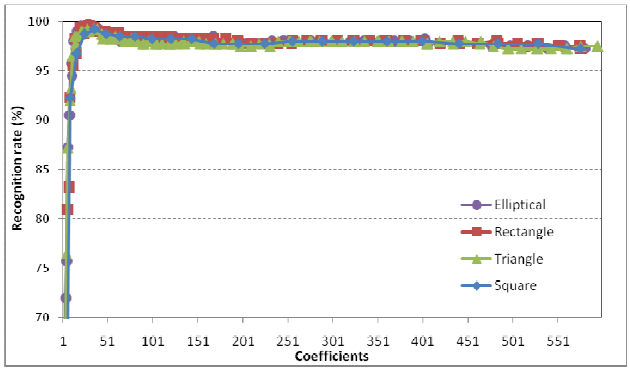


Figure 9: Effect of varying the number of DCT coefficients on recognition accuracy, using geometric regions.

In Figure 10, the curve apex region is highlighted. In this region lies the most significant results. It can be noted that the elliptical format has the best result: only 29 DCT coefficients are enough to achieve an accuracy of 99.75%. This rate represents only one error in 400 classifications. This result was obtained using an elliptical form with a size of 7x6 pixels. If the ORL database was larger, covering more people also would require a larger amount of coefficients for a correct distinction between the classes, due to the dimensionality curse.

The square region always crosses the rectangular region, since a rectangular region can also be a square one. The rectangular region has hit rates greater than the square of the region, reaching 99.50%, while the square only reaches 99.25%. However, the square region was shown for comparison with the previous work of Matos [11], which used only square regions. Using elliptical forms, the detection rate was 99.75%. Comparing to the Matos' work, the error rate was reduced from 0.75% to only 0.25%. This was a 66% substantial reduction in the error rate.

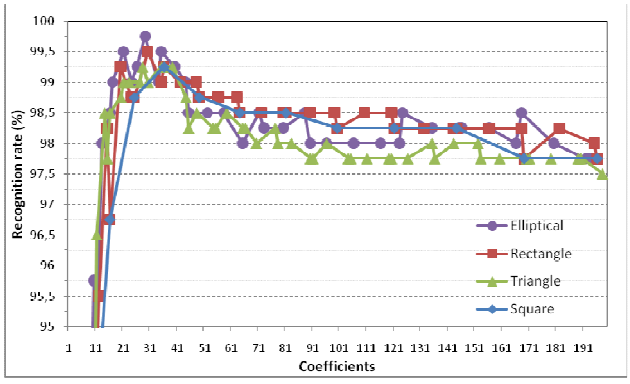


Figure 10: Zoom at the region with best results from Figure 9.

B. Cumulative leave-one-out

Using this approach, the system reaches 100% of recognition accuracy when the correct face is within a range of six returned faces. This range of six faces was obtained with both rectangular and triangular regions. The elliptical region reached 100% only within seven faces returned. This result was not the expected, because the elliptical regions have been more efficient under the common leave-one-out approach. The square region reached the same result only within 9 faces. Figure 11 shows these results.

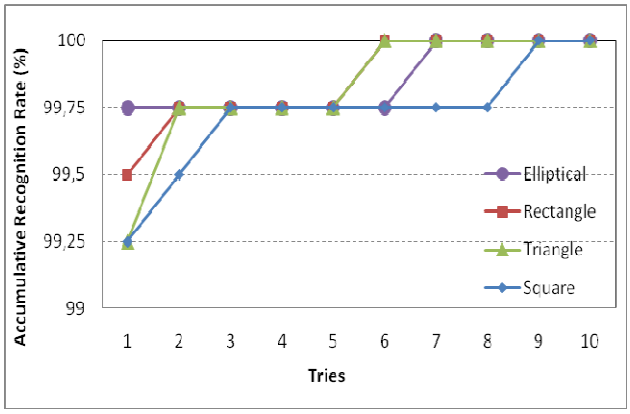


Figure 11: Accumulative Recognition

It is interesting to see that all the four shapes reached a recognition rate higher than 99% when returning only the first face. The results with just one returned face can also be perceived in Figure 10, by selecting only the best results for each geometric shape.

C. Automatic zooming normalization

Using the Automatic Zoom Normalization approach on the ORL database, a recognition rate of 100% was achieved for all the geometric shapes used in the attribute selection stage. This shows that the zooming normalization is a very important step in order to make the system more robust to small scale variations. This sensitivity to zoom is due to the intrinsic properties of the DCT. However, by applying a zoom in and out of only 5% on a face database where the faces size are slightly normalized, it is sufficient to overcome the sensitivity of the DCT.

Table 1 shows the regions, their size, and the number of coefficients used to achieve 100% of recognition. It can be noted that, within very small regions, it is possible to obtain 100%. In the common leave-one-out approach, the best result was reached by using elliptical regions of size 6x7, leading to only 29 coefficients. Using the same approach, but now applying the Zooming Normalization procedure, the system achieves the optimal result with regions of size 3x4, with only 12 coefficients.

Shape	Size	coeffients
square	4x4	16
rectangle	3x4	12
triangular	4x5	14
elliptical	4x5	13

Table 1: Shape, size and coefficients that hit 100%.

D. Processing time

The Face Recognition System proposed in this paper was tested on a 2.0 GHz AMD Turion64 single core computer with 1GB of RAM, running Ubuntu Linux, a common PC of nowadays. Using this platform, the execution time with two different implementations was verified: one developed in Java and the other developed in C++, both implemented during the development of this work.

Most of the execution time is spent with the generation of the DCTs of the face database – approximately 99% of the execution time. However, this step can be performed only once and its result stored for a future use. Table 2 shows the execution time of the leave-one-out test with the ORL database using the C++ implementation, and Table 3 shows the execution time using the Java implementation. The second column of these tables shows the execution time when the DCTs are fully generated, and the third column shows the time when the DCTs were pre-computed and are only loaded from the disk. The leave-one-out round (second row) has 400 classifications, and the time to classify only one face (third row) is the time of the leave-one-out round divided by 400.

	Execution time (seconds)	
	Generating DCTs	Loading DCTs
Obtain DCTs	15.660030	0.076801
Leave-one-out round	0.119613	0.119613
1 face classification	0.000299	0.000299
Total time	15.779643	0.196414

Table 2: C++ execution time.

	Execution time (seconds)	
	Generating DCTs	Loading DCTs
Obtain DCTs	42,400000	0,789000
Leave-one-out round	0,202000	0,202000
1 face classification	0,000505	0,000505
Total time	42,602000	0,991000

Table 3: Java execution time.

It can been seen that the C++ version is faster in all aspects, and it is 10 times faster on doing the DCT loading operation. Thus, the C++ implementation is more suitable for real-time applications.

VI. DISCUSSION AND CONCLUSIONS

The proposed face recognition method has proved to be valid in the performed tests, achieving good results in several evaluation categories as recognition accuracy, robustness and computational cost.

Without any pre-processing step, the proposed method achieves a high recognition accuracy of 99.75% when using only 29 coefficients using the leave-one-out approach. On the cumulative leave-one-out the system reaches 100% of correct classification when the six most similar faces are returned. When applying the automatic zooming normalization the system also reaches 100% of correct detection rate.

The proposed method is also suitable for real time applications: in the experimental tests, the classification processing time for only one face, using 29 coefficients, is near 0.0002 seconds, in a common PC, without taking into account the prior feature selection processing time. Due to energy compaction property of DCT it is possible to reduce the processing of 10.304 (112x92) pixels to 29 DCT coefficients.

Testing the proposed algorithm using only the ORL database is not enough to fully validate the method, but it's important to compare methods. Despite the fact that in literature this face database is always being referred, it does not include all variations needed for evaluating the robustness of the method.

Future developments to the method will focus on improving the processing time on very large databases, and integrate the system with a face detector method. Additionally, tests must be done on other face databases.

VII. REFERENCES

[1] AT&T Laboratories, Cambridge, UK, "The ORL Database of Faces" (Now At&T "The Database of Faces"), Available [Online]: http://www.cl.cam.ac.uk/research/dtg/attarchive/pub/data/att_faces.zip [November, 16, 2008], 1994.

[2] Batista, L. V. Compressão de Sinais Eletrocardiográficos Baseada na Transformada Cosseno Discreta. D.Sc. Thesis in Electrical Engineering, PPGEE /DEE/UFPB, Campina Grande, Brazil, 2002.

[3] Bicego, M., Castellani, U, And Murino V, Using HMM and Wavelets for Face Recognition, Proceedings of the 12th International Conference on Image Analysis and Processing, 2003 IEEE.

[4] Campos, T. E., Técnicas de Seleção de Características com aplicação em Reconhecimento de Faces, Dissertação de Mestrado, USP, São Paulo, Brasil, 2001.

[5] Duda, R. O.; Hart, P. E.; and Stork, D. G. Pattern Classification. John Wiley & Sons, Inc., 2nd edition, 2003.

[6] Ekenel, H. K.; Goa, S. H.; Fischerm, M.; and Stiefelhagen, R. Face Recognition for Smart Interactions, IEEE ICME, 2007.
[7] Gonzalez, Rafael C.; Woods, Richard R.. Digital Image Processing Using Matlab. 3ED. PRENTICE HALL, 2007.
[8] Haykin, Simon. Neural Networks and Learning Machines. 3rd Edition. Prentice Hall, 2008.
[9] Hafed, Ziad M and Levine, Marin D. Face Recognition Using Discrete Cosine Transform. International Journal of Computer Vision, v. 43(3), p. 167-188. 2001.
[10] Kumar, S, A, S; Deepti, D, R, And Prabhakar, B, Face Recognition Using Pseudo-2d Ergodic HMM, IEEE, ICASSP 2006.
[11] Matos, F. M. de S. Reconhecimento de Faces Utilizando Seleção de Coeficientes da Transformada Cosseno Discreta. M.Sc. Dissertation in Computer Science, PPGI/DI/UFPB, João Pessoa, Brazil, 2008.
[12] Podilchuk, C. and Zhang, X. 1996. Face recognition using DCT-based feature vectors. In Proceedings of the Acoustics, Speech, and Signal Processing, 1996. on Conference Proceedings., 1996 IEEE international Conference – Volume 04 (May 07 - 10, 1996). ICASSP. IEEE Computer Society, Washington, DC, 2144-2147.
[13] Rao, k. R,; YIP, P.. Discrete Cosine Transform: Algorithms, Advantages, Applications. Academic Press, Inc, 1990.
[14] Zhao, W.; Chellappa, R.; Phillips, P. J., Rosenfeld, A.; Face Recognition: A Literature Survey, ACM Computing Surveys, V, 35, N, 4, P, 399-458, 2003.