

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Visage - Impacto dos Filtros de Abstração no Reconhecimento Facial em Imagens

Pedro Tiago Pontes

DISSERTAÇÃO PROVISÓRIA



Mestrado Integrado em Engenharia Informática e Computação

Orientador: Cristina Ribeiro

Co-orientador: Filipe Coelho

16 de Junho de 2013

Visage - Impacto dos Filtros de Abstração no Reconhecimento Facial em Imagens

Pedro Tiago Pontes

Mestrado Integrado em Engenharia Informática e Computação

Resumo

O reconhecimento facial é efetuado de uma forma rotineira e quase sem esforço por parte das pessoas no seu dia-a-dia. No entanto, a construção de sistemas automáticos de reconhecimento facial é uma tarefa complexa que engloba um conjunto de sub-problemas específicos, nomeadamente, a deteção e segmentação das faces presentes na imagem, a sua normalização ao nível da forma e iluminação e a extração das características distintivas das faces, para que, por fim, seja efetuado o reconhecimento da identidade das pessoas representadas. À resolução com sucesso deste conjunto de sub-problemas, colocam-se também um conjunto de desafios dos quais se destacam a variação ao nível da pose, iluminação e expressão das pessoas representadas.

Os desafios inerentes ao processo de reconhecimento, assim como, a vasta gama de aplicações onde a identificação de indivíduos é necessária, como por exemplo o controlo de acesso a informação, o reforço da segurança, o reforço da aplicação da lei, o entretenimento e a gestão de conteúdos multimédia, despoletou a atenção de inúmeros investigadores ao longo dos últimos 40 anos. Como consequência, verificou-se uma evolução notável ao nível da eficácia dos sistemas desenvolvidos, tal como pode ser verificado pelo registo de uma melhoria de duas ordens de grandeza desde 1993 nas últimas avaliações efetuadas. Em cenários cooperativos com condições de captura de imagens controladas, nomeadamente ao nível da pose, iluminação e expressões faciais, considera-se mesmo que o problema de reconhecimento facial se encontra praticamente resolvido. Por outro lado, em cenários não cooperativos esta é ainda uma área de investigação em aberto. Para isso, contribui também o facto de o grau de satisfação dos utilizadores dos sistemas estar diretamente relacionado com a sua área de aplicação e com os dispositivos através dos quais é efetuado o reconhecimento, sendo que, diferentes áreas de aplicação e diferentes dispositivos exigem soluções diferentes e portanto investigação diferenciada.

Os filtros de abstração constituem uma forma moderna de simplificação do conteúdo visual, permitindo remover informação redundante e dar apenas destaque à mensagem visual a transmitir. O projeto Visage propõe o desenvolvimento de um sistema de reconhecimento facial de personalidades, onde seja utilizada a abstração de imagens no processo de reconhecimento, com o objetivo de analisar o impacto dos filtros de abstração nesse processo. Para além do impacto dos filtros de abstração pretende-se também analisar qual a importância de um conjunto de tarefas de pré-processamento efetuadas sobre as imagens no desempenho do sistema criado.

O sistema desenvolvido tem por base bibliotecas de código aberto que implementam algoritmos de reconhecimento facial próximos do estado da arte e o uso do filtro de abstração Kuwahara Anisotrópico. Com base neste sistema será possível desenvolver aplicações de pesquisa visual capazes de encontrar fotos de uma personalidade específica ou apresentar, para uma foto de rosto fornecida pelo utilizador, a celebridade ou figura pública mais parecida.

Abstract

Face recognition is performed in a routine and effortless way by people in their daily life. However, building automated face recognition systems is a complex task that comprises a set of specific sub-problems, namely, face detection and segmentation, normalization of the extracted face in terms of light and shape and the extraction of the distinguishing features of the face, so that the recognition of the person present in a picture can be performed. In order to solve each of sub-problems with success, a set of challenges must be overcome, from which variation in pose, illumination and facial expression variation can be highlighted.

The challenges associated with the recognition process, as well as, the wide range of areas where the identification of people is required, such as information access control, security, law reinforcement, entertainment and multimedia content management, raised the attention of numerous researchers in the past 40 years. Consequently, a remarkable evolution was noticed in terms of system's performance, as can be seen by the two orders of magnitude improvement since 1993, registered in the last evaluation performed. In cooperative scenarios with controlled image capturing conditions, it is considered that the problem of face recognition is almost resolved. By the other hand, in non-cooperative scenarios this is still an open research area. Nevertheless, the user satisfaction level of a system is directly related to its application area and the device with which recognition is being performed, whereas different areas of application and different devices require distinguished research.

Abstraction filters are a modern way to simplify visual content, allowing the removal of useless detail and the use of abstraction for effective communication. Visage project proposes the development of a face recognition system of public figures, where image abstraction is used in the process of recognition, with the aim of analyzing the impact of abstraction filters in face recognition. Furthermore, the impact in the performance of the system of a set of image pre-processing steps were also analyzed in the aim of this project.

The system developed uses open source libraries that implement face recognition algorithms near to the state of the art and uses Kuwahara Anisotropic filter for image abstraction. With this system, the development of visual retrieval applications, capable of finding photos of public figures or, when a photo is provided by the user, finding the most similar celebrity to a person will be made available.

Conteúdo

1	Introdução	1
1.1	Motivação	2
1.2	Objetivos	2
1.3	Estrutura do Relatório	3
2	Reconhecimento Facial em Imagens	4
2.1	Evolução	4
2.2	Definição do Problema	5
2.2.1	Sub-Problemas	6
2.2.2	Desafios	7
2.3	Áreas de Aplicação	9
2.3.1	Controlo de Acesso e Segurança de Informação	9
2.3.2	Cartões Inteligentes e Identificação de Faces	10
2.3.3	Entretenimento	11
2.3.4	Gestão de Conteúdos Multimédia e Bases de Dados	11
2.3.5	Segurança e Aplicação da Lei	12
2.4	Estratégias de Reconhecimento Facial	12
2.4.1	Soluções Existentes	15
2.5	Análises de Desempenho Efetuadas	16
2.5.1	FERET	17
2.5.2	FRVT	18
3	Abstração de Imagens	22
3.1	Estilização Artística de Imagens	22
3.2	Filtros de Abstração de Imagens	23
3.2.1	Filtro Gaussiano	24
3.2.2	Filtro Bilateral	24
3.2.3	Kuwahara	25
3.2.4	<i>Diffusion Shock Filter</i>	26
3.2.5	Filtros Morfológicos	26
3.2.6	Técnicas <i>Gradient Domain</i>	26
3.3	Trabalho Relacionado	27
4	Sistema de Reconhecimento Facial Visage	28
4.1	Filtros de Abstração de Imagens no Reconhecimento Facial	28
4.2	Visage - Visão geral	29
4.3	Coleção de dados - <i>Labeled Faces in the Wild</i>	30
4.4	Arquitetura	33

CONTEÚDO

4.5	Módulo <i>FaceDetector</i>	33
4.5.1	Deteção Face	34
4.5.2	Alinhamento e Corte	37
4.5.3	Normalização Contraste	39
4.5.4	Aplicação Filtro Abstração	40
4.5.5	Aplicação Máscara	41
4.6	Módulo <i>FaceModel</i>	41
4.6.1	<i>Eigenfaces</i>	42
4.6.2	<i>Fisherfaces</i>	43
4.6.3	<i>Local Binary Patterns Histograms (LBPH)</i>	43
4.6.4	Top <i>N</i> Resultados	44
4.7	Funcionalidades	45
5	Avaliação e Resultados	47
5.1	Conjuntos de Teste	47
5.2	Pré-processamento	50
5.2.1	Deteção e segmentação da face	50
5.2.2	Normalização do Contraste	51
5.2.3	Abstração Imagens	52
5.3	Avaliação <i>Closed-Set Identification</i>	52
5.3.1	Metodologia Avaliação	53
5.3.2	Resultados Experiência 1: Impacto da segmentação das faces	53
5.3.3	Resultados Experiência 2: Impacto da Normalização do Contraste	56
5.3.4	Resultados Experiência 3: Abstração de Imagens	59
5.4	Avaliação <i>Image Retrieval</i>	61
5.4.1	Metodologia Avaliação	62
5.4.2	Resultados	63
5.5	Variação Desempenho	64
5.6	Análise Necessidade Armazenamento	66
5.7	Discussão Resultados Obtidos	66
6	Conclusões e Perspetivas Futuras	67
6.1	Conclusões	67
6.2	Perspectivas Futuras	68
	Referências	69

Lista de Figuras

2.1	Sistema Genérico de Reconhecimento Facial	8
2.2	A redução da taxa de erro para os algoritmos estados da arte tal como documentado nas avaliações FERET, FRVT 2002, FRVT 2006 e MBE 2010 [PSO07, GQP10]	20
3.1	Cronologia do desenvolvimento da área da IB-AR [KCWI12].	23
3.2	Diferentes filtros de abstração aplicados sobre a mesma imagem [KKD09].	24
3.3	<i>Diffusion Shock Filter</i> [KL08].	26
4.1	Distribuição do número de imagens por pessoa na biblioteca LFW.	31
4.2	Diferentes versões da biblioteca LFW	32
4.3	Diferentes tarefas de pré-processamento disponíveis no sistema Visage.	34
4.4	Comparação entre face correctamente detectada e situações de erro na deteção	35
4.5	Comparação entre versões original, recortada, alinhada e LFW-a.	38
4.6	Comparação entre imagem original e as várias técnicas de normalização disponíveis.	39
4.7	Comparação entre imagem original e os vários filtros de abstração utilizados.	40
4.8	Exemplo de face já segmentada, com e sem máscara.	41
5.1	Exemplo de imagens do conjunto \mathcal{G} para o sujeito Angelina Jolie	49
5.2	Exemplo de imagens do conjunto \mathcal{P} para o sujeito Angelina Jolie	49
5.3	Variação da mesma imagem nos vários conjuntos pré-processados analisados.	51
5.4	Desempenho algoritmos Eigenfaces, Fisherfaces e LBPH para os conjuntos Original, Cropped e Masked	54
5.5	Comparação do desempenho dos três algoritmos implementados para o conjunto Masked	56
5.6	Desempenho algoritmos Eigenfaces, Fisherfaces e LBPH para os conjuntos Normalized, Equalized e CLAHE	57
5.7	Comparação do desempenho dos três algoritmos implementados para o conjunto Equalized	58
5.8	Desempenho algoritmos Eigenfaces, Fisherfaces e LBPH para os conjuntos Gaussian, Bilateral e AKF	60
5.9	Comparação do desempenho dos três algoritmos implementados para o conjunto AKF	61

Lista de Tabelas

2.1	Categorização dos Métodos de Reconhecimento Facial em Imagens por Zhao <i>et al.</i> [ZCPR03]	14
2.2	Participantes na avaliação MBE 2010 [GQP10].	16
4.1	Caracterização da Biblioteca LFW	32
4.2	Resultados etapa de de detecção facial para todas as pessoas com pelo menos 2 imagens na biblioteca LFW-a.	36
5.1	Galerias criadas após pré-processamento.	50
5.2	Resultados da Média da Precisão na Galeria (melhores a negrito).	63
5.3	Algoritmos ordenados por desempenho nos diversos conjuntos de teste (L = LBPH, F = Fisherfaces, E = Eigenfaces; melhores a negrito).	64
5.4	Desvio padrão (σ) entre os valores de MPG obtidos para os 4 conjuntos de teste (melhores a negrito).	65
5.5	Necessidades de armazenamento galerias pré-processadas.	66

Abreviaturas e Símbolos

ACP	Análise dos Componentes Principais
ALD	Análise Linear Discriminante
API	<i>Application Programming Interface</i> (Interface de Programação de Aplicações)
CPU	<i>Central Processing Unit</i> (Unidade central de processamento)
LBP	<i>Local Binary Patterns</i>
LBPH	<i>Local Binary Patterns Histograms</i>
LFW	<i>Labeled Faces in the Wild</i>
FERET	<i>Face Recognition Technology</i>
FRVT	<i>Facial Recognition Vendor Test</i>
GPU	<i>Graphics Processing Unit</i> (Unidade de Processamento Gráfico)
HCInt	<i>High Computational Intesity test</i>
MBE	<i>Multiple Biometric Evaluation</i>
MCF	<i>Mean Curvature Flow</i>
MCInt	<i>Medium Computational Intensity test</i>
NIST	<i>National Institute of Standards and Technology</i>
OpenCV	<i>Open Source Computer Vision Library</i>
RAPID	Reconhecimento Automático de Passageiros Identificados Documentalmente
ROC	receiver operating characteristic

Capítulo 1

Introdução

A sociedade atual caracteriza-se pelo constante fluxo de informação a que os seus indivíduos estão expostos, assistindo-se a um crescimento acelerado na produção de conteúdos multimédia, principalmente tendo em conta os meios de transmissão de informação digital [Int]. O aparecimento de novas formas de comunicação, como por exemplo as redes sociais, e uso de um maior leque de dispositivos para a produção de conteúdos como por exemplo *smartphones* e *tablets*, intensificou ainda mais este crescimento, contribuindo também para o surgimento de conteúdos cada vez mais personalizados e variados. Neste contexto, torna-se importante o desenvolvimento de novas metodologias que possibilitem a categorização, organização e posterior pesquisa por parte dos utilizadores destes conteúdos, representando o reconhecimento facial automático em imagens um papel importante a esse nível.

O reconhecimento facial é um componente importante da capacidade de perceção humana, sendo efetuado de uma forma rotineira e quase sem esforço por parte dos seres humanos. Contudo, a construção de sistemas computacionais capazes de efetuar este tipo de reconhecimento de uma forma semelhante aos humanos é ainda uma área de investigação em aberto [LJ11]. Este é um problema de tal forma desafiante e interessante que tem despertado a atenção de investigadores das mais diversas áreas, ao longo dos últimos 40 anos, tais como: psicologia, reconhecimento de padrões, redes neuronais, visão por computador e computação gráfica [ZCPR03].

Por outro lado, a abstração de imagens dota os artistas de novas ferramentas de transmissão de informação, tendo como objetivo, melhorar eficácia da comunicação visual. Os filtros de abstração permitem remover informação não essencial de uma imagem, dando apenas destaque à mensagem a transmitir.

No âmbito desta dissertação, é levantada a seguinte hipótese: O uso de abstração em imagens que vão ser posteriormente alvo de reconhecimento facial, pode melhorar o processo de reconhecimento. Este impacto deverá ser analisado do ponto de vista da eficácia do reconhecimento, mas também ao nível das necessidades de processamento e armazenamento das coleções de dados a reconhecer.

1.1 Motivação

O reconhecimento facial automático em imagens é uma área de investigação em aberto, principalmente quando considerados ambientes onde as condições de captura de imagens não são controladas, uma situação muito comum na grande maioria das bibliotecas de imagens existentes. Nessas situações, as grandes variações existentes ao nível da iluminação, pose e expressão dos indivíduos propõem grandes desafios à tarefa de reconhecimento facial automático para os quais existe ainda a necessidade de evoluir as soluções existentes.

A pesquisa ao nível do reconhecimento facial é motivada não só pelos desafios inerentes ao processo de reconhecimento facial, mas também, pelas inúmeras aplicações onde a identificação de indivíduos é necessária [LJ11]. A esse nível, a evolução tecnológica registada nos últimos vinte anos culminou com o surgimento de uma panóplia de dispositivos e áreas onde se verifica um aumento do interesse na aplicação desta tecnologia, desde a mais tradicional segurança e controlo de informação, a áreas como o entretenimento ou a gestão de conteúdos multimédia e bases de dados. Ao nível dos dispositivos, o surgimento de novas plataformas, e o aumento da capacidade de computação das já existentes, torna possível a aplicação destes sistemas, não só em máquinas dedicadas para o efeito, mas também em computadores pessoais ou mesmo em dispositivos móveis como *smartphones* ou *tablets*.

Por outro lado, o elevado valor comercial associado aos sistemas de reconhecimento facial existentes, tem como consequência a existência de um número reduzido de soluções abertas, fazendo com que os resultados eventualmente obtidos no âmbito desta dissertação possam ter uma boa visibilidade.

Finalmente, a aplicação de filtros de abstração de imagens para a ilustração automática de texto demonstrou melhorias ao nível da informação retornada, assim como, na redução significativa das necessidades de processamento e armazenamento necessárias [CR12], tornando-se assim pertinente o estudo do impacto dos filtros de abstração no processo de reconhecimento facial no âmbito desta dissertação.

1.2 Objetivos

O principal objetivo desta dissertação visa o desenvolvimento de um sistema de reconhecimento facial automático, que permita o estudo do impacto da utilização de filtros de abstração visual de informação, assim como de várias etapas de pré-processamento, no processo de reconhecimento facial automático em imagens. Para isso, destacam-se o seguinte conjunto de objetivos parciais:

1. Desenvolvimento de um sistema de reconhecimento facial de personalidades, baseado em software livre;
2. Integração de diferentes tarefas de pré-processamento de imagens de modo a aumentar a robustez do sistema;

3. Integração da abstração de imagens no sistema desenvolvido;
4. Avaliação do impacto das diferentes tarefas de pré-processamento, assim como da abstração de imagens no processo de reconhecimento facial;

O sistema desenvolvido deverá ser capaz de detetar as faces presentes numa imagem e apresentar uma lista de possíveis entidades nela contidas, permitindo, desta forma, quer a realização avaliação levada a cabo no âmbito desta dissertação, quer a construção de uma base sólida para o desenvolvimento de futuras aplicações que tiram partido do reconhecimento facial automático em imagens no seu funcionamento.

1.3 Estrutura do Relatório

Este relatório encontra-se dividido em seis capítulos. No primeiro capítulo deste relatório, após uma breve introdução, são sintetizados os objetivos e motivação desta dissertação.

De seguida, é apresentada uma revisão do problema de reconhecimento facial automático em imagens, analisando os desafios existentes, as suas diversas áreas de aplicação e as estratégias adotadas para os sistemas atualmente existentes. No terceiro capítulo deste relatório, é abordado o tema da abstração de imagens, ilustrando as soluções existentes, assim como o trabalho relacionado com esta dissertação efetuado com recurso à abstração de imagens.

Em quarto lugar, encontra-se descrito o sistema de reconhecimento facial desenvolvido, a sua arquitetura, os seus principais módulos e funcionalidades e ainda a biblioteca de imagens utilizada para o desenvolvimento e avaliação do sistema criado.

O quinto capítulo retrata os temas avaliação e resultados. Nele são apresentados os conjuntos de teste criados para a avaliação do sistema Visage, as diferentes galerias criadas após pré-processamento das imagens, as metodologias de avaliação adotadas e os seus respetivos resultados.

Por último, no sexto capítulo, encontram-se sintetizadas as principais conclusões a retirar do trabalho desenvolvido, assim como, algum do trabalho futuro a realizar.

Capítulo 2

Reconhecimento Facial em Imagens

Neste capítulo é efetuada, em primeiro lugar, uma introdução ao problema de reconhecimento facial automático em imagens através de uma apresentação da sua evolução histórica. Posteriormente, é apresentada a definição do problema de reconhecimento facial automático, assim como dos sub-problemas que o compõem e dos desafios existentes para a sua resolução. Em terceiro lugar, são apresentadas as diversas áreas de aplicação do reconhecimento facial automático. De seguida, é efetuada uma revisão das estratégias utilizadas para a implementação deste tipo de sistemas, assim como das diversas soluções atualmente existentes. Por último, é efetuada uma revisão das avaliações efetuadas no âmbito dos programas FERET e FRVT e são apresentadas as conclusões retiradas destas avaliações.

2.1 Evolução

A capacidade de distinguir indivíduos através das suas características faciais é inata aos seres humanos, efetuada de uma forma natural e quase sem esforço. A forma como efetuamos esse reconhecimento, é no entanto um problema complexo, que atrai a atenção de investigadores das mais diversas áreas, sendo da psicologia e na década de 50 [BT54], o mais antigo estudo efetuado sobre o tema. Associado a este problema, o reconhecimento facial automático, ou seja, o reconhecimento facial efetuado de uma forma automática com o auxílio de máquinas tais como computadores, tem de igual forma recebido muita atenção por parte da comunidade científica, sendo que o primeiro sistema de reconhecimento facial automático de que há registo, foi desenvolvido na década de 70 por Takeo Kanade [Kan73]. Ao longo dos anos, investigadores das mais diversas áreas, tais como psicologia, reconhecimento de padrões, neurologia, visão por computador ou computação gráfica, efetuaram estudos em vários aspetos do reconhecimento facial efetuado por computadores e máquinas.

Ao nível do reconhecimento facial automático, após o trabalho de Kanade, e outros realizados por alguns dos seus pares na década de 70, verificou-se um abrandamento na investigação científica efetuada na área até meados dos anos 90, onde o trabalho realizado por Turk e Pentland

[TP91] trouxe um novo impulso ao desenvolvimento de sistemas e metodologias de reconhecimento facial automático. Neste estudo, Turk e Pentland apresentaram uma nova forma abordagem ao reconhecimento facial designada de *Eigenfaces*. Um outro marco no desenvolvimento nas metodologias de reconhecimento facial, foi o surgimento do método *Fisherfaces* [BHK97, ZCK98], o qual apresenta uma maior insensibilidade em relação à variação da iluminação e expressões faciais presentes nas imagens.

Ao longo da década de 90 a investigação científica na área do reconhecimento facial automático intensificou-se, tendência que ainda se verifica atualmente, tal como pode ser verificado pelo surgimento de várias conferências, das quais são exemplo, desde as mais antigas *International Conference on Automatic Face and Gesture Recognition (AFGR)* (1995) e *International Conference on Audio- and Video-Based Authentication (AVBPA)* (1997), até à recentemente realizada *5th International Conference on Biometrics (ICBS)* (2012).

Paralelamente, o surgimento de programas como o Face Recognition Technology (FERET) (1993-1998), levado a cabo pelo *National Institute of Standards and Technology (NIST)* dos E.U.A, com o objetivo de financiar a investigação na área e estabelecer padrões para a avaliação das diversas metodologias de reconhecimento facial criadas, assim como impulsionar a criação de uma base de dados de faces de grande dimensões para o teste dos sistemas criados, permitiu uma evolução significativa na área [PWH98, PRR00].

No final da década de 90, surgiram as primeiras aplicações comerciais de reconhecimento facial automático, tal como pode ser comprovado pela participação de 5 soluções comerciais na primeira edição dos *Facial Recognition Vendor Test (FRVT)*, no ano 2000 [BBP01]. Nesta avaliação, as soluções eram testadas nas componentes de usabilidade e performance do reconhecimento efetuado, de modo fornecer uma avaliação objetiva e independente dos sistemas de reconhecimento facial presentes no mercado. Estas avaliações foram repetidas nos anos de 2002, 2006, 2010 e 2012.

2.2 Definição do Problema

De uma forma geral, o problema de reconhecimento facial pode ser formulado da seguinte forma: dada uma imagem (prova), pretende-se identificar ou verificar a presença de uma ou mais pessoas na prova, comparando-a com uma base de dados de faces previamente guardada (galeria) [YKMA02, ZCPR03]. Este problema, é ainda geralmente dividido em três subproblemas específicos: verificação de faces (ou autenticação), *pair matching* e identificação de faces [LJ11, HRBLm07]:

Verificação Uma prova é apresentada ao sistema, assim como a respetiva identidade da pessoa presente na imagem. Cabe ao sistema verificar se existe uma correspondência entre a pessoa representada na imagem e a identidade fornecida. Assim, é efetuada uma comparação 1:1, que responde à pergunta: "Esta pessoa é quem realmente afirma ser?".

Pair Matching São apresentadas duas provas ao sistema, cada uma contendo uma face. O sistema deve decidir se as duas imagens correspondem, ou não, a representações da mesma pessoa. Tal como no sub-problema de verificação é efetuada uma comparação 1:1, a qual responde à pergunta: "Estas imagens pertencem à mesma pessoa?".

Identificação Uma prova é apresentada ao sistema, e pretende-se que este devolva a identidade da pessoa presente na imagem, assim como o nível de confiança associado à correspondência efetuada entre a imagem e a identidade devolvida. A identificação de faces pode ainda ser dividida em dois sub-problemas:

No primeiro problema designado *watch list* ou *open-set identification* [CSP10], não é garantido que o sistema conheça a pessoa a ser identificada, pelo que em primeiro lugar é necessário identificar se a pessoa existe na galeria e, caso exista, determinar a identidade da pessoa. Este sub-problema responde às perguntas: "Esta pessoa está presente na galeria? Se sim, de quem é esta face?".

O segundo caso, é aquele que é geralmente avaliado em sistemas de reconhecimento facial, onde é garantido que a pessoa é conhecida pelo sistema, respondendo-se apenas à pergunta: "De quem é esta face?". Nesta situação, designada de *closed-set identification*, podem ainda ser pedidas todas as correspondências efetuadas com um nível de confiança superior a um valor pré-definido.

Em qualquer um dos subproblemas de identificação, é efetuada uma comparação 1:N.

Os sistemas de reconhecimento facial podem ainda ser divididos genericamente em dois grupos, tendo em conta a fonte de onde são extraídas as faces a analisar. O primeiro grupo baseia-se no uso de imagens estáticas, como por exemplo fotografias de personalidades ou mesmo em retratos de faces humanas criadas por ilustradores. O segundo, tem por base o uso de vídeos, dos quais são extraídas as faces a identificar ou verificar, sendo que neste caso essa identificação pode ainda ser efetuada em tempo real ou *à posteriori*. No entanto, cada um destes meios representa problemas e desafios diferentes ao reconhecimento facial automático, pelo que devem ser estudados de forma individualizada, sendo que, no âmbito desta dissertação irá apenas ser abordado o sub-problema de reconhecimento facial efetuado em imagens estáticas.

2.2.1 Sub-Problemas

O problema de reconhecimento facial em imagens é um problema complexo que obriga a que um conjunto de sub-tarefas sejam efetuadas com sucesso para uma identificação ou verificação eficaz da pessoa presente na imagem. Em primeiro lugar, é necessário determinar onde se encontra a face da pessoa a ser identificada, separando-a do resto da imagem. Posteriormente é necessário encontrar características que diferenciem essa face de outras. Finalmente, é ainda necessário efetuar a comparação da face, e respetivas características, com a galeria de imagens existente, fazendo assim o reconhecimento da pessoa representada. Este conjunto de sub-tarefas, são geralmente designados pelos seguintes passos fundamentais [ZCPR03]:

Deteção: Determina a presença de uma face na imagem, e em caso positivo, efetua a sua segmentação do resto da imagem.

Extração características faciais: Consiste na extração de características úteis para a diferenciação entre as diversas faces presentes no sistema. As características extraídas variam dependendo do método utilizado, podendo ir desde a distância entre olhos, à textura e cor da pele da pessoa a ser reconhecida, por exemplo.

Identificação ou Verificação Consiste na identificação ou verificação da identidade do sujeito presente na imagem tendo em conta as características extraídas e a informação presente no sistema.

Para além dos três passos mencionados anteriormente, existe ainda um quarto passo, efetuado após a deteção e antes da extração das características faciais, que merece destaque em alguns dos sistemas de reconhecimento facial existentes [LJ11]:

Normalização Introduzida em alguns métodos de reconhecimento facial com o objetivo de melhorar os resultados do reconhecimento ao nível pose e iluminação das faces nas imagens. O seu objetivo consiste em normalizar a face geometricamente, ou seja na sua forma e enquadramento, e fotometricamente, ou seja ao nível da iluminação.

Na figura 2.1 é possível ver uma representação dos quatro passos fundamentais enunciados anteriormente que constituem um sistema de reconhecimento facial genérico. Em algumas soluções os passos de deteção e extração de características possuem tarefas em comum, sendo realizados simultaneamente.

2.2.2 Desafios

O reconhecimento facial automático em imagens é um problema complexo, sendo que existem um conjunto de desafios que se colocam na construção de um sistema de tipo. O primeiro passo em qualquer um desses sistemas, é a deteção de faces na imagem, a esse nível Yang *et al.*, destacaram, no revisão ao estado da arte realizado em 2002, os seguintes desafios [YKMA02]:

- **Pose e Orientação:** A posição relativa da face à câmara pode sofrer uma grande variação dependendo do ambiente onde é capturada a imagem (variação no ângulo da face em relação à câmara vertical e horizontalmente, posição frontal, perfil). Ao variar a posição da face, alguns dos elementos que a permitem localizar e reconhecer podem ficar obstruídos. Avaliações efetuadas [BBP01], demonstram que este impacto é maior, quanto maior for o ângulo da pose em relação à posição frontal, sendo que uma variação até cerca de 25° não afeta significativamente os resultados do reconhecimento.
- **Elementos Estruturais:** a presença ou ausência de elementos estruturais, tais como barba ou óculos, assim como, a grande variação de forma e cor a que estes elementos podem apresentar afetam significativamente a forma como a face é percebida pelo sistema.

Reconhecimento Facial em Imagens

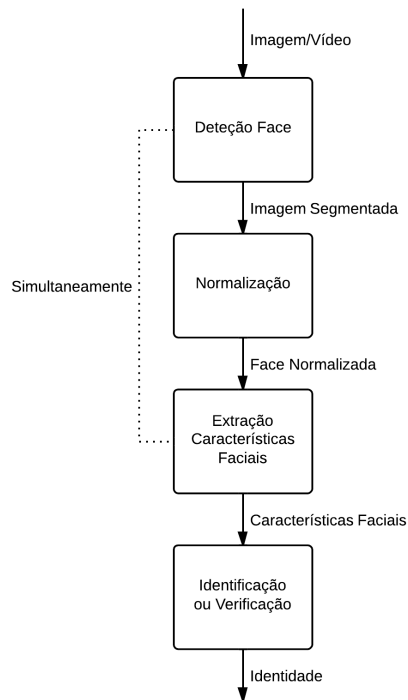


Figura 2.1: Sistema Genérico de Reconhecimento Facial

- Obstrução da face: as imagens captadas podem possuir objetos em frente à cara da pessoa a ser identificada, levando apenas a uma representação parcial da face na imagem.
- Condições da imagem: aspetos como a resolução da imagem, o sensor utilizado para a captura e até mesmo eventuais defeitos associados à câmara utilizada podem potencialmente afetar o aspeto final da face na imagem.

Outro fator que afeta significativamente a perceção de uma face por um sistema de reconhecimento facial automático é a iluminação. Uma imagem pode ser capturada, com iluminação natural ou artificial. Para além disso, a iluminação pode estar sujeita a uma grande variação de tonalidades, intensidade e ângulo de incidência da luz.

Mesmo quando captadas pela mesma câmara, com as mesmas condições de iluminação e no mesmo local, podem existir variações significativas nas imagens captadas da face de uma pessoa devido às diferentes expressões faciais contidas nas imagens, fazendo com que este seja outro desafio a ultrapassar ao efetuar o reconhecimento facial automático.

O envelhecimento, ou eventuais alterações provocadas na face devido a fatores externos como por exemplo acidentes, exposição a condições extremas ou intervenção cirúrgica é outro dos fatores a ter em conta na identificação de uma pessoa.

Finalmente, existem ainda fatores como o elevado grau de semelhança entre pessoas com graus de parentesco próximos. Imagens capturadas de gémeos, ou até mesmo de pais e filhos, podem

à primeira vista parecer imagens da mesma pessoa, tornando ainda mais difícil para um sistema automático efetuar a distinção dos indivíduos.

Assim sendo, para um reconhecimento fiável e eficaz é essencial que todas as etapas identificadas em 2.2.1 sejam concluídas com sucesso. Sem uma correta deteção da presença de uma face na imagem, todos os passos subjacentes seriam impossíveis de ser realizados, por outro lado, o sucesso da extração das características faciais, encontram-se também dependente de uma normalização eficaz das imagens. No entanto, apesar desta dependência verificada entre as diversas etapas, cada uma delas representa problemas complexos, que devem a merecer a atenção e estudo de forma individual para que uma maior evolução das diversas técnicas envolvidas seja atingida [ZCPR03].

2.3 Áreas de Aplicação

O problema de reconhecimento facial automático tem merecido a atenção e estudo de inúmeros investigadores ao longo dos últimos 40 anos, motivados não só pelos desafios inerentes ao processo de reconhecimento facial, mas também pelas pela vasta gama aplicações onde a identificação de indivíduos é necessária [LJ11].

Adicionalmente, a evolução tecnológica da última década, permitiu a criação de sistemas computacionais mais poderosos, os quais tornaram possíveis desenvolvimentos na área que anteriormente seriam impensáveis gerando um ainda da maior interesse, tal como pode ser verificado pelo surgimento de uma vasta gama de aplicações comerciais e governamentais de reconhecimento facial, como é exemplo o sistema de fronteira automática dos aeroportos portugueses [Minb], assim como, pela recente aquisição de empresas da área do reconhecimento facial por parte de grandes empresas tecnológicas, nomeadamente Google e Facebook.

De seguida encontram-se descritas, em cinco categorias, algumas das áreas de aplicação do reconhecimento facial automático, respetivamente ilustradas com exemplos de aplicação real. Esta categorização foi baseada no estudo efetuado em [LJ11], o qual é para nosso conhecimento o mais recente estudo efetuado às diversas áreas de aplicação do reconhecimento facial, assim como em estudos anteriores efetuados por [ZCPR03]. Esta secção pretende apenas apresentar de uma forma sumária algumas das possíveis áreas de aplicação, e respetivas soluções, às quais o reconhecimento facial poderá ser aplicado, não pretendendo ser um estudo exaustivo de todas as soluções existentes.

2.3.1 Controlo de Acesso e Segurança de Informação

A necessidade de sistemas de fácil utilização que assegurem a segurança da nossa informação digital, assim como da privacidade dos utilizadores, sem que para isso seja necessário a memorização e utilização de um sem número de palavras-chave e códigos de segurança é premente. Neste campo, o reconhecimento facial destaca-se ao permitir a criação de sistemas de autenticação não invasivos que requerem um nível de cooperação baixo por parte dos participantes, em contraste

com os sistemas tradicionais como o uso de palavras-chave, ou mesmo de sistemas mais complexos como a análise de impressões digitais, retina ou íris. Por outro lado, este tipo de sistemas, permite ainda determinar a presença factual do utilizador que se pretende identificar, e não apenas determinar o conhecimento de um código de acesso à informação.

Atualmente, o uso de reconhecimento facial para o controlo de acesso a informação encontra-se já a ser utilizado em diversas aplicações comerciais como são exemplos os sistemas de autenticação disponíveis numa grande variedade de computadores portáteis. De uma forma geral, nestes sistemas, o utilizador apenas necessita de se aproximar do computador para que o reconhecimento facial e consequente autenticação sejam efetuados. Desta forma o utilizador não necessita de efetuar qualquer intervenção para aceder aos seus dados ou de memorizar qualquer palavra-chave.

O uso de reconhecimento facial para garantir uma maior segurança de informação, ou efetuar um mais eficaz controlo do seu acesso, é geralmente aplicado em soluções com um número de utilizadores limitado e nas quais as imagens capturadas apresentam baixa variabilidade nas condições de iluminação e pose dos utilizadores. Desta forma, o nível de precisão do reconhecimento atingido nesta área é normalmente elevado, fazendo com que o nível de satisfação dos seus utilizadores seja também elevado [LJ11].

2.3.2 Cartões Inteligentes e Identificação de Faces

Os documentos pessoais inteligentes registam uma presença cada vez mais ativa na sociedade atual, como é caso do cartão do cidadão português (CC), que à data de 5 de Novembro de 2012 já tinha sido adotado por mais de 7 milhões de portugueses [paMaA], ou mesmo do passaporte eletrónico português (PEP), o qual recebeu igualmente uma ampla aceitação desde a sua criação em 2006 [Mina].

Tradicionalmente, este tipo de documentos encontram-se dotados de capacidade de processamento e armazenamento próprios, podendo interagir com aplicações de terceiros como, por exemplo, sistemas de autenticação com recurso ao reconhecimento facial. O sistema designado de Reconhecimento Automático de Passageiros Identificados Documentalmente (RAPID)[Minb], desenvolvido pela empresa portuguesa Vision-Box [VB] encontra-se desde 2007 em funcionamento nos vários aeroportos internacionais portugueses e tira partido das vantagens da associação da tecnologia de reconhecimento facial com documentos inteligentes. Em primeiro lugar, o sistema verifica se os dados armazenados no chip do passaporte são válidos, posteriormente, é efetuada uma comparação entre a fotografia armazenada no cartão e a imagem capturada do passageiro pelo dispositivo instalado na fronteira automática. Caso a identificação seja verificada, a porta que permite a passagem entre fronteiras é então automaticamente aberta. Todo o processo, encontra-se ainda a ser supervisionado por um operador numa cabine, o qual se encontra a supervisionar múltiplos dispositivos de identificação, sendo apenas necessária a sua intervenção quando não é efetuada uma identificação válida da pessoa.

O nível de precisão dos sistemas que tiram partido das sinergias entre cartões inteligentes e o reconhecimento facial apresenta uma correlação com o nível de atualização dos dados presentes nos documentos, sendo que com um nível de atualização frequente é possível ser atingido um nível

de satisfação elevado [LJ11], como se comprova pela bom funcionamento do sistema RAPID. Contudo, esta é ainda uma área de investigação em aberto para que uma utilização destes sistemas de forma totalmente autónoma (sem que seja necessária a supervisão de um operador), em larga escala e, por exemplo, em ambientes não controlados seja possível.

2.3.3 Entretenimento

As tecnologias de reconhecimento facial apresentam uma forte potencialidade de aplicação nas áreas associadas ao entretenimento e à produção de conteúdos digitais, nomeadamente no que diz respeito à produção de jogos de computador e na interação humano computador.

As consolas modernas estão dotadas de capacidade de processamento notáveis, para além disso, juntamente com estas consolas é cada vez mais comum o surgimento de dispositivos como o *Kinect* da Microsoft que permitem novas formas de interação e a criação de novos tipos de jogos. Uma das novidades introduzidas pelo *Kinect* é a inclusão de diversos sensores e câmaras que permitem, entre outras coisas, efetuar um reconhecimento facial dos utilizadores. Desta forma, as equipas de desenvolvimento de jogos de computador, podem agora de uma forma mais fácil utilizar mecanismos de reconhecimento facial automático nos seus jogos de forma a criar jogos mais imersivos e personalizados. Este reconhecimento pode ainda ser aplicado ao nível da interação pessoa-computador onde permite, a criação de sistemas com uma maior facilidade de interação, uma vez que o próprio rosto poderá servir como um dispositivo de *input* natural.

À semelhança dos sistemas de controlo de acesso o reconhecimento facial automático na área do entretenimento é geralmente aplicado em soluções com um número de utilizadores limitado e onde o nível de precisão necessário não é muito alto fazendo assim com que a satisfação dos seus utilizadores seja, em geral, elevada [LJ11].

2.3.4 Gestão de Conteúdos Multimédia e Bases de Dados

A sociedade atual caracteriza-se pelo constante fluxo de informação a que os seus indivíduos se encontram expostos. A título de exemplo e de acordo com dados divulgados pela Intel[Int], a cada minuto, são visualizados no Youtube 1,3 milhões de vídeos e carregadas para os seus servidores 30 novas horas de vídeo. Já o Flickr, no mesmo período de tempo, regista a visualização de 20 milhões de fotos e o carregamento de 3000 novas imagens para a sua plataforma. Segundo os mesmos dados, atualmente o número de dispositivos com ligação à Internet é equivalente ao da população mundial e é esperado que em 2015 seja o dobro dessa população.

Este crescimento verificado quer na quantidade de conteúdos produzidos, quer na diversidade de dispositivos que acedem a esses conteúdos, torna assim extremamente importante uma gestão e organização eficazes dos mesmos. Os métodos tradicionais de pesquisa e organização de conteúdos multimédia baseiam-se em anotações textuais associadas a estes conteúdos. Com o uso de sistemas de reconhecimento facial é possível efetuar o reconhecimento das entidades presentes numa biblioteca multimédia mesmo que as suas fotografias não possuam qualquer anotação

textual, uma vez que é feita uma análise da própria imagem e não apenas dos seus descritores, tornando-se assim um importante auxílio dos métodos tradicionais.

Atualmente existem no mercado diversas aplicações que permitem efetuar a organização de álbuns fotográficos pessoais com recurso ao reconhecimento facial como por exemplo o Picasa da Google ou o iPhoto da Apple. Por outro lado, redes sociais como Google Plus e o Facebook, lançaram também recentemente serviços que permitem aos seus utilizadores a deteção e identificação automática de outros utilizadores presentes nas imagens carregadas para os seus servidores.

Apesar de a performance destes sistemas ter registado uma evolução notável nos últimos anos, a grande variabilidade associada às condições de captura de imagens e a grande quantidade de dados a ser processada faz com que exista ainda uma grande margem de evolução possível a este nível, sendo que a performance verificada ao nível do reconhecimento efetuado varia muito de acordo com a aplicação a analisar, assim como as características do conjunto de dados a organizar.

2.3.5 Segurança e Aplicação da Lei

A preocupação com a segurança de instalações públicas e privadas, o controlo de emigração ilegal, a segurança de eventos com um grande impacto mediático ou mesmo o auxílio à investigação policial tornaram a segurança e aplicação da lei uma das áreas de aplicação que mais interesse despoletou por parte de sectores privados e governamentais desde o surgimento das primeiras tecnologias de reconhecimento facial.

Uma das aplicações mais comuns do reconhecimento facial em segurança é o reforço da vigilância da via pública. Desde 1998, diversas cidades implementaram sistemas de vigilância com recurso a reconhecimento facial, como são exemplos os sistemas instalados em *Newham Borough of London*, *Tampa (Florida)* e *Virginia Beach (Virginia)* [LJ11]. Um outro exemplo, foi o sistema desenvolvido para os jogos olímpicos de 2008 em Pequim pela *Chinese Academy Of Sciences* e que foi utilizado para validar as entradas nas cerimónias de abertura e encerramento dos jogos olímpicos [Chi].

A aplicação de sistemas de reconhecimento facial para a segurança e aplicação da lei tem atraído muita atenção por parte de sectores privados e governamentais, havendo o registo de alguns casos onde a sua aplicação se revelou um sucesso. Contudo, de uma forma geral estes sistemas debatem-se com um baixo nível de satisfação por parte dos seus utilizadores, isto porque, existe uma grande variabilidade nas condições de iluminação das imagens captadas, pose dos utilizadores, assim como, um número muito elevado de faces a analisar, resultando num número elevado de falsos positivos e consequentemente numa má performance dos sistemas [LJ11].

2.4 Estratégias de Reconhecimento Facial

Após o trabalho inicial de Takeo Kanade em 1973 [Kan73], a área do reconhecimento facial automático passou um período de dormência até meados da década de 90, onde os trabalhos realizados recorrendo a métodos como Análise dos Componentes Principais (ACP) (*Principal Component Analysis*), Análise Linear Discriminante (ALD) (*Linear Discriminant analysis*) e Elastic

Graph Matching (EGM), revitalizaram a área [CSP10]. Ao nível da ACP, o método *Eingenfaces* desenvolvido por Turk e Pentland em 1991 [TP91], foi um marco fundamental no rejuvenescimento do reconhecimento facial automático. Este método tira partido da redundância natural existente entre as representações faciais de diversos indivíduos, para através de uma análise dos componentes principais das faces, desenvolvida anteriormente por Kirby e Sirovich [KS90], efetuar uma representação de baixo nível das faces existentes. Posteriormente, o método *Fisherfaces* [BHK97, EC97, ZCK98], que tira partido da aplicação de ALD após uma análise inicial dos componentes principais da imagem de forma a maximizar as variações entre indivíduos, apresentou também resultados positivos em experiências em bases de dados com um número elevado de faces a analisar. Finalmente, ao nível de EGM, foi precursor o trabalho realizado por Wiskott *et al.* [WFKvdM97]. No algoritmo EGM as faces são representadas como grafos, sendo cada nó posicionado em pontos fiduciais da face e a cada aresta associado um vetor de distância. A identificação de uma face consiste em determinar entre os grafos existentes aquele que maximiza a função de similaridade entre grafos.

Desde então, diversos investigadores estenderam estes três tipos de algoritmos. Em 2003, Zhao *et al.* [ZCPR03] efetuaram um estudo aprofundado das técnicas desenvolvidas até então classificando-as em três categorias. Métodos holísticos, caso estes usem toda a região da face como fonte de comparação no reconhecimento, métodos baseados nas características faciais, caso estes usem as características extraídas como informação essencial para a classificação da face e ainda como híbridos, caso seja utilizada uma mistura de ambos os métodos. Os resultados desta classificação podem ser visualizados na tabela 2.1.

Tal como referido na secção 2.2.2 deste relatório, o reconhecimento facial é uma tarefa complexa, composta por um conjunto de sub-problemas que enfrentam uma série de desafios para a sua conclusão. O primeiro desafio em qualquer sistema de reconhecimento facial é a deteção da face na imagem, a esse nível o trabalho realizado por Paul Viola e Michael Jones [VJ04] é considerado como um dos mais robustos métodos disponíveis na atualidade. Outras áreas onde a investigação atual se encontra focada têm a ver com a criação de métodos de reconhecimento facial robustos em relação à iluminação das imagens, assim como a pose dos indivíduos representados. [CSP10].

Ao nível da pose, o uso de *3D morphable face models* [BV03] tem revelado resultados positivos em posições não frontais. Estes algoritmos simulam o processo de formação de uma face em 3D, estimando a sua representação tridimensional com recurso a técnicas de computação gráfica, a partir de imagens 2D fornecidas ao sistema. Esta estimativa é efetuada através do encaixe de um modelo 3D de faces na própria imagem fornecida, modelo esse apreendido através da digitalização 3D de um conjunto de rostos e respetiva textura. No entanto, em grande parte dos modelos desenvolvidos a seleção manual de um pequeno número de características faciais é requerida e o poder de computação necessário para a criação dos modelos é elevado. Extensões destes modelos foram também propostas para criar sistemas mais robustos em relação à variação na iluminação [CSP10].

Tabela 2.1: Categorização dos Métodos de Reconhecimento Facial em Imagens por Zhao *et al.* [ZCPR03]

Método	Trabalho Representativo
Holistic methods	
<i>Principal Component analysis (PCA)</i>	
Eigenfaces	Direct application of PCA [Craw and Cameron 1996; Kirby and Sirovich 1990; Turk and Pentland 1991]
Probabilistic eigenfaces	Two-class problem with prob. measure [Moghaddam and Pentland 1997]
Fisherfaces/subspace LDA	FLD on eigenspace [Belhumeur et al. 1997; Swets and Weng 1996b; Zhao et al. 1998]
SVM	Two-class problem based on SVM [Phillips 1998]
Evolution pursuit	Enhanced GA learning [Liu and Wechsler 2000a]
Feature lines	Point-to-line distance based [Li and Lu 1999]
ICA	ICA-based feature analysis [Bartlett et al. 1998]
<i>Other Representations</i>	
LDA/FLD	LDA/FLD on raw image [Etemad and Chellappa 1997]
PDBNN	Probabilistic decision based NN [Lin et al. 1997]
Feature-based methods	
Pure geometry methods	Earlier methods [Kanade 1973; Kelly 1970]; recent methods [Cox et al. 1996; Manjunath et al. 1992]
Dynamic link architecture	Graph matching methods [Okada et al. 1998; Wiskott et al. 1997]
Hidden Markov model	HMM methods [Nefian and Hayes 1998; Samaria 1994; Samaria and Young 1994]
Hybrid methods	
Convolution Neural Network	SOM learning based CNN methods [Lawrence et al. 1997]
Modular eigenfaces	Eigenfaces and eigenmodules [Pentland et al. 1994]
Hybrid LFA	Local feature method [Penev and Atick 1996]
Shape-normalized	Flexible appearance models [Lanitis et al. 1995]
Component-based	Face region and components [Huang et al. 2003]

Outras abordagens que têm em vista a criação de sistemas de reconhecimento facial robustos em relação à iluminação centram os seus esforços na estimação do albedo para a normalização das imagens. O albedo representa a relação entre a quantidade de luz refletida por um ponto e quantidade de luz incidente, métodos recentes centram-se no desenvolvimento de filtros estocásticos não estacionários para a estimação de mapas de albedo das amostras faciais [BAC09]. Estes mapas podem ainda ser combinados com os modelos 3D faciais desenvolvidos para uma representação facial mais completa. Apesar de os resultados obtidos por estes métodos serem promissores quando comparados com os métodos tradicionais como o *Eigenfaces*, estes métodos ainda carecem de maior validação, uma vez que as suas avaliações foram feitas utilizando conjuntos de dados controlados e de reduzida dimensão [CSP10].

O investimento realizado na investigação de metodologias de reconhecimento facial automático tem permitido progressos notáveis na área. Do mesmo modo, a investigação em áreas relacionadas como a deteção de faces nas imagens, o envelhecimento facial ou o reconhecimento de expressões faciais em imagens tem permitido evoluir ainda mais os resultados obtidos. Devido à complexidade envolvida no processo de reconhecimento facial automático a evolução na área encontra-se assim intrinsecamente ligada à evolução verificada nas múltiplas áreas envolvidas, sendo que a investigação independente de cada uma dessas áreas é crítica para uma resolução eficaz da grande variabilidade de desafios que o reconhecimento facial automático se propõe ultrapassar.

2.4.1 Soluções Existentes

A investigação desenvolvida nos vários centros de investigação na década de 90, levou à passagem das soluções de reconhecimento facial automático de meros protótipos laboratoriais, para soluções comerciais de elevado valor acrescentado, dessas soluções resultaram três categorias comerciais de produtos na área do reconhecimento facial:

- **Soluções Completas.** Estas soluções caracterizam-se por oferecer toda a infraestrutura necessária para efetuar o reconhecimento facial, quer a nível de software, quer a nível de hardware.
- **Software.** Soluções tradicionais de sistemas de reconhecimento facial, que consistem em software que permite efetuar o reconhecimento facial independentemente do hardware utilizado para a captura de informação. Dentro destas soluções existe uma grande variedade nos produtos oferecidos, desde soluções baseadas na captura de imagens 2D estáticas, imagens 3D e vídeo. Para além disso, as soluções oferecidas encontram disponíveis para serem utilizadas em diversas plataformas, como por exemplo dispositivos móveis ou através da Internet.
- **Software Development Kit (SDK)** de reconhecimento facial. Permitem aos clientes, a criação de novas soluções de reconhecimento facial a partir de módulos independentes ou APIs desenvolvidas por terceiros.

Tabela 2.2: Participantes na avaliação MBE 2010 [GQP10].

Empresa/Organização	Website
Cognitec	http://www.cognitec-systems.de
Dalian University of Technology	http://www.dlut.edu.cn
L1 Identity Solutions	http://www.l1id.com ¹
NEC	http://www.nec.com/en/global/solutions/security/technologies/face_recognition.html
Neurotechnology	http://www.neurotechnology.com
Pittsburg Pattern Recognition	http://www.pittpatt.com ²
Sagem	http://www.sagem.com ³
Surrey University	http://www.surrey.ac.uk
Toshiba	http://www.toshiba.com
Tsinghua University	http://www.tsinghua.edu.cn

Para além do reconhecimento facial automático, as soluções disponíveis possuem muitas vezes também associados outros métodos de biometria, como por exemplo reconhecimento de íris ou impressões digitais, assim como métodos mais tradicionais de autenticação como o uso de cartões de identificação ou passwords, de modo a aumentar a robustez e segurança dos sistemas. Na tabela 2.2, encontram-se listadas as soluções participantes na última avaliação, para a qual existem dados disponíveis, realizada a sistemas de reconhecimento facial automático em imagens.

Para além das soluções comerciais de reconhecimento facial ao longo dos últimos anos surgiram também algumas soluções gratuitas de reconhecimento facial. Deste tipo de soluções são exemplo o software Picasa da Google, ou software iPhoto da Apple que permitem aos seus utilizadores a organização de álbuns multimédia com recurso ao reconhecimento facial. Por outro lado, redes sociais como Google Plus e o Facebook, lançaram também recentemente serviços que permitem a deteção e identificação automática de utilizadores presentes nas imagens carregadas para os seus servidores. Finalmente ao nível das plataformas de desenvolvimento, existem também alguns sistemas de código aberto, sendo que a este nível se destaca a biblioteca *OpenCV*, pelo seu grau de maturidade, grande número de utilizadores e dimensão da comunidade de suporte.

2.5 Análises de Desempenho Efetuadas

O desempenho de um sistema de reconhecimento facial depende de uma grande variedade de fatores como a iluminação, posição da face na imagem, expressão facial e presença ou ausência de adereços (ex:óculos) na face. Por outro lado, as condições de utilização do próprio sistema, tais como o número de utilizadores diferentes e a frequência de utilização, são também fatores

¹Integrada no grupo Morphotrust USA - <http://www.morphotrust.com>

²Adquirida pela Google Inc.

³Integrada no grupo Morphotrust USA - <http://www.morphotrust.com>

a ter em conta ao avaliar a performance de um sistema. Uma avaliação apropriada dos sistemas de reconhecimento facial encontra-se então dependente da existência de uma base de dados de grandes dimensões e consequentemente um conjunto de casos de teste vasto, assim como, da existência de métodos apropriados para essa avaliação.

A criação do programa FERET em 1993, pelo NIST, permitiu assim colmatar uma lacuna existente na avaliação dos sistemas de reconhecimento facial existentes com a criação de uma base de dados de faces de grande dimensões assim como de um protocolo de avaliação dos respetivos sistemas de reconhecimento facial. Com base no protocolo desenvolvido foram levadas a cabo avaliações aos anos de 1994 e 1995. Um novo protocolo foi criado para a avaliação realizada em 1996. Posteriormente, o mesmo instituto realizou novas avaliações baseadas no protocolo FERET96 na edição de 2000 dos FRVT, a partir de 2002, um outro protocolo foi criado (protocolo FRVT 2002), com base no protocolo FERET96, mas com a particularidade de permitir a avaliação de sistemas de biometria em geral e não apenas sistemas de reconhecimento facial. Com base no novo protocolo criado foram então realizadas avaliações nos anos de 2002, 2006 e 2010. A decorrer encontram-se ainda as avaliações FRVT 2012, para as quais ainda não existem à data resultados divulgados.

As avaliações levadas a cabo durante os programas FERET e FRVT constituem desta forma uma boa base para a análise da evolução dos sistemas de reconhecimento facial automático ao longo dos últimos 20 anos, tal como pode ser visto na figura 2.2. De seguida encontram-se descritas os principais objetivos e conclusões retiradas dos programas FERET e FRVT:

2.5.1 FERET

Para além dos objetivos anteriormente citados de criação de uma base de dados de grandes dimensões e de criação de um protocolo que permitisse uma avaliação concreta e de base científica dos algoritmos existentes, o programa FERET tinha ainda como propósitos avaliar o estado da arte e identificar futuras áreas de desenvolvimento do reconhecimento facial [PRR00]. Tendo em conta estes objetivos, foi então avaliada a performance dos diferentes algoritmos em diferentes cenários, categorias de imagens e diferentes versões dos próprios algoritmos. A performance foi computada para situações de identificação e verificação de faces, estando os principais resultados da última avaliação, efetuada em 1996, descritos em Phillips *et al.* [PRR00].

A primeira avaliação do programa FERET, em 1994, consistia num três de conjuntos de testes, cada um com uma galeria e conjunto de provas diferentes. O primeiro conjunto, consistia numa prova de reconhecimento de faces de uma galeria de 316 faces. O segundo, tinha com objetivo determinar a capacidade dos algoritmos de rejeitar faces não presentes na galeria, designadas de falso-alarme. O terceiro, tinha como objetivo analisar os efeitos da pose na performance dos algoritmos [PRR00].

Em 1995, teve lugar a segunda avaliação, com o objetivo de avaliar os mesmo algoritmos em galerias de maior dimensão. A performance foi medida num único teste, com uma galeria de 817 indivíduos conhecidos, tendo sido dado um maior ênfase na avaliação de provas duplicadas. Um

duplicado é uma imagem cuja imagem correspondente da galeria foi geralmente capturada em dias diferentes [PRR00].

Na última edição, em 1996, e já com um protocolo atualizado, foram testados 12 algoritmos de reconhecimento facial, e estudada a sua performance em quatro situações diferentes, usando para isso uma galeria com um total de 1196 indivíduos. Na primeira situação, a galeria e as imagens de prova foram tiradas no mesmo dia e sobre as mesmas condições de iluminação. No segundo conjunto de provas, a galeria e as imagens de prova foram tiradas em dias diferentes. Para o terceiro teste, a galeria e as respectivas provas foram capturadas com pelo menos um ano de diferença. Finalmente, no quarto e último conjunto de testes, a galeria e as imagens de prova foram tiradas no mesmo dia, mas com diferentes condições de iluminação.

Várias conclusões foram retiradas das avaliações efetuadas. A primeira conclusão foi que, de facto, se verificava uma evolução nas técnicas de reconhecimento facial, uma vez que os resultados foram progressivamente melhores, mesmo quando comparadas versões diferentes do mesmo algoritmo. Essa conclusão é suportada pela análise dos resultados do caso geral de identificação, testado em todas as edições, onde as imagens da galeria e da prova foram capturadas no mesmo dia e com a mesma iluminação. Nesse teste em 1994, apenas 78% indivíduos foram identificados com sucesso numa galeria de 317 indivíduos, ao passo que em 1995 foram identificados já 93 % dos indivíduos numa galeria de 831 e nos testes da última edição (1996), 95% dos indivíduos foi identificado numa galeria de 1196 indivíduos. De igual forma, foi também verificada evolução na deteção de duplicados, mantendo-se no entanto as percentagens de identificação de duplicados mais reduzidas quando comparadas com o caso geral de identificação devido à maior complexidade do problema.

Outra conclusão retirada foi que a performance se encontra dependente da galeria e conjunto de imagens de prova utilizadas, uma vez que foi verificado que mudando a galeria, mas mantendo o mesmo tipo de teste a ser efetuado a percentagem de indivíduos identificados varia significativamente. Para a galeria e conjunto de prova capturadas no mesmo dia a percentagem varia entre os 80% e os 94%, enquanto que para galeria e conjunto de prova, capturadas em dias diferentes essa percentagem varia entre os 24% e os 69% por cento. Para além disso, não se existiu qualquer algoritmo que se revelasse superior em todas as quatro situações de teste realizadas em 1996, variando as performances, conforme a situação de teste.

Finalmente, conforme a tarefa a efetuar, identificação ou verificação de faces, ficou também demonstrado que não havia uma técnica que se superiorizasse em ambas as situações, comprovando que o reconhecimento facial é uma tarefa complexa que necessita de um estudo e investigação concretos tendo em conta a área de aplicação em vista.

2.5.2 FRVT

Durante a década de 90 registou-se um progresso assinalável nas diversas técnicas de reconhecimento facial, tendo para isso sido muito importante o papel desempenhado por iniciativas como o FERET. No final da década de 90, assistiu-se então a uma passagem dos sistemas de reconhecimento facial existentes de meros protótipos universitários, para aplicações comerciais de

reconhecimento facial automático, tal como pode ser comprovado pela participação de 5 soluções comerciais na primeira edição dos *Facial Recognition Vendor Test*, no ano 2000. Esta avaliação, tinham como objetivo maioritário efetuar uma avaliação técnica das capacidades dos sistemas comerciais de reconhecimento facial disponíveis, de forma a compreender os pontos fortes e fracos de cada sistema, obtendo também uma análise do estado da arte do reconhecimento facial à data [BBP01]. Estas avaliações foram ainda repetidas em 2002, 2006, 2010 e ainda em 2012.

A primeira edição dos FRVT, em 2000, utilizou o mesmo protocolo de avaliação da avaliação última avaliação FERET em 1996, no entanto, os testes efetuados eram significativamente mais exigentes do que a avaliação anterior, sendo utilizada também uma muito maior variedade de imagens. Os resultados da avaliação efetuada foram ser reportados para as seguintes oito categorias: compressão, distância, expressão, media, iluminação, pose, resolução e variação temporal. Ao nível das conclusões retiradas, podem ser destacados os seguintes tópicos [BBP01, CSP10, LJ11]:

- Compressão de imagens, utilizado JPEG, até 40:1 não reduziu a taxa de reconhecimento;
- Pose não afeta o reconhecimento de forma significativa até $\pm 25^\circ$, mas afeta significativamente quando são atingidos os $\pm 40^\circ$;
- De forma equivalente ao registado no programa FERET, o uso de diferentes iluminações interiores não afeta significativamente os resultados obtidos, no entanto esses resultados são afetados caso haja uma mudança entre zonas interiores e exteriores;
- Registou-se uma melhoria significativa dos resultados obtidos em imagens tiradas com mais de 18 meses de distância, quando comparado com os resultados obtidos no FERET. (Aumento de 5% a 12% de faces identificadas conforme o conjunto de dados avaliado).

A segunda edição dos FRVT teve lugar em 2002, contou com a participação de soluções de reconhecimento facial de 10 empresas distintas. Esta edição consistiu em duas provas fundamentais, os *High Computational Intesity test (HCInt)* e os *Medium Computational Intensity test (MCInt)*. Os HCInt tinham como objetivo medir a performance dos sistemas existentes em 121 589 imagens frontais de 37 435 pessoas, representando um desafio computacional de elevado grau de exigência, como comprovado pelas 242 horas de computação necessárias para efetuar os testes. Os *Medium Computational Intensity test (MCInt)*, assemelhavam-se às medições efectuadas em avaliações anteriores, como o FERET e o FRVT 2000, tendo como objetivo medir a performance dos sistemas em imagens de diferentes categorias, nomeadamente não frontais, capturas no interior e exterior, entre outros. As conclusões retiradas desta avaliação foram as seguintes [CSP10, LJ11]:

- Melhores sistemas são insensíveis a variações de luz interior;
- Reconhecimento facial em imagens exteriores é ainda um desafio por resolver;
- Utilização de *3D morphable models* melhora a performance em situações não frontais;
- Características como sexo e idade podem afetar significativamente a performance, sendo que se registou maior facilidade em distinguir homens e pessoas com idade superior.

Reconhecimento Facial em Imagens

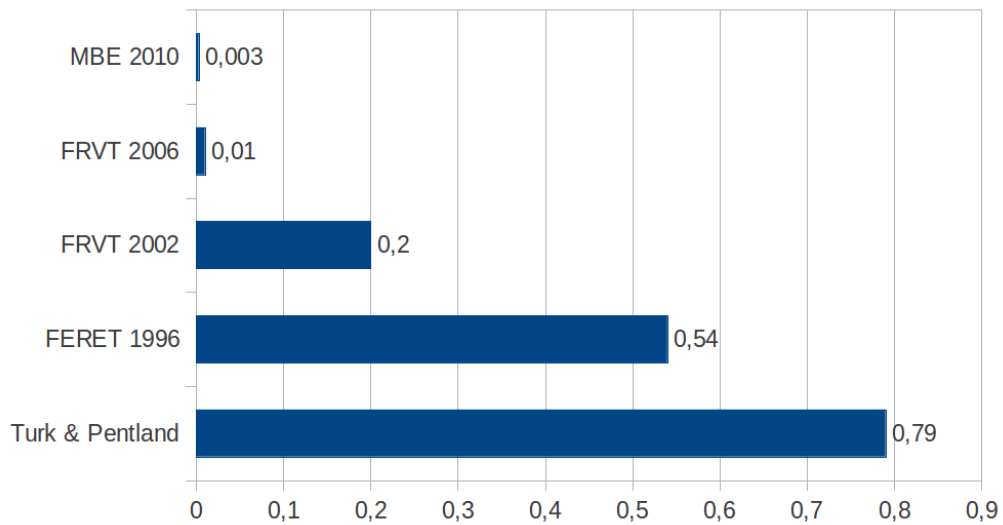


Figura 2.2: A redução da taxa de erro para os algoritmos estados da arte tal como documentado nas avaliações FERET, FRVT 2002, FRVT 2006 e MBE 2010 [PSO07, GQP10]

- O tempo de captura entre provas afeta significativamente a performance, registrando-se uma diminuição linear da performance com o aumento do tempo.

As últimas edições dos FRVT para as quais foram divulgados resultados foram as edições de 2006 [PSO07] e a de 2010. Na edição de 2010 a competição não assumiu o nome de FRVT 2010, mas sim de *Multiple Biometric Evaluation (MBE) 2010-still image track*, uma vez que além de avaliados sistemas de reconhecimento facial automático foram também avaliados sistemas de biometria com base na análise da íris. [GQP10]. Foi também realizada uma edição dos FRTV em 2012, mas a avaliação dos sistemas ainda se encontra a decorrer, não havendo por isso resultados divulgados.

Em 2006, foi analisada a performance de sistemas de reconhecimento facial, de 22 organizações e de 10 países diferentes, baseados em representações 3D e em imagens estáticas de faces. As avaliações consistiam em três testes distintos: comparação de imagens capturadas com luz interior; comparação de digitalizações de faces 3D com base na forma e textura; comparação de imagens capturadas em estúdio com imagens capturadas em corredores e átrios. Em 2010, para além do conjunto de dados analisados em 2006 e 2002, de modo a averiguar a evolução comparativa dos algoritmos, foi também analisado um segundo conjunto de imagens recolhidas por várias agências de segurança e compiladas pelo *Federal Bureau of Investigation (FBI)*. Este segundo conjunto de dados possuía mais de 1 milhão de faces fazendo da avaliação de 2010 a avaliação em maior escala feita aos sistemas de reconhecimento facial à data. Das provas de 2006 e 2010, as principais conclusões obtidas foram as seguintes [PSO07, GQP10, CSP10, LJ11]:

- Desde 1993 foram atingidas melhorias de 2 ordens de magnitude;

Reconhecimento Facial em Imagens

- Os resultados de 2010 e 2002 mostram uma melhoria de 1 ordem de magnitude nos sistemas de reconhecimento facial (Para um *false accept rate (FAR)* de 0.001, foi registrado em 2002 um *false reject rate (FRR)* de 0.2, enquanto que em 2010 $FRR=0.038$);
- Em condições de captura de informação controlada, a performance medida de sistemas de reconhecimento facial e sistemas baseados na análise da íris foi equivalente;
- A performance de sistemas baseados em imagens estáticas e sistemas baseados em representações 3D da face foi equivalente;
- Iluminação e resolução das imagens representam um papel fundamental no reconhecimento facial;

Capítulo 3

Abstração de Imagens

Neste capítulo é efetuada uma análise das diversas abordagens existentes para a abstração de imagens. Inicialmente, é apresentada uma contextualização da abstração de imagens, no âmbito da estilização artística de imagens, onde são abordadas de forma sumária as diferentes técnicas existentes para o efeito. De seguida, encontram-se descritos alguns dos filtros de abstração de imagens mais comuns. Por último, é ainda apresentado um trabalho relacionado com esta dissertação na área de abstração de imagens.

3.1 Estilização Artística de Imagens

Os conteúdos foto-realistas, possuem muitas vezes mais informação do que a necessária para uma comunicação visual eficaz. A área da Estilização Artística de Imagens (*Image Based Artistic Rendering*), onde se insere a abstração de imagens, tem como objetivo fazer a transformação de imagens e vídeos foto-realistas 2D em novas formas de representação visual simplificada. A estilização de imagens e vídeos permite, para além da simplificação do conteúdo visual, a sua estruturação de forma a dar ênfase e enfoque a determinadas características das imagens. Desta forma, é possível guiar a perceção obtida dos conteúdos visuais conforme o objetivo da mensagem a comunicar. Por outro lado, estas técnicas permitem criar novas formas de representação gráfica de informação, dotando assim os artistas de novas ferramentas de comunicação e criação de conteúdos visuais.

Atualmente a área da estilização artística de imagens diversificou-se numa atividade multidisciplinar que engloba áreas como a visão por computador, interação humano computador, computação gráfica e a modelação perceptual (*perceptual modeling*). Os sistemas de estilização evoluíram também desde sistemas semi-automáticos de processamento de imagem no início dos anos 90, até sistemas totalmente automáticos de renderização. Adicionalmente o recurso a técnicas de visão por computador e o uso de filtros de abstração de imagens, permitiram melhorias notáveis ao nível da manipulação de grandes quantidades de informação, assim como da informação visual produzida. Uma evolução desta técnicas até atualidade pode ser vista na imagem [3.1](#).

Abstração de Imagens

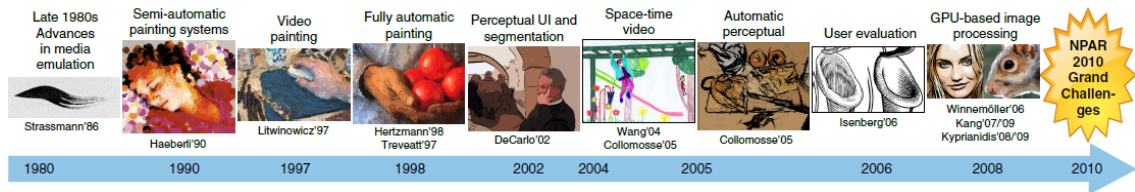


Figura 3.1: Cronologia do desenvolvimento da área da IB-AR [KCWI12].

De acordo com o estudo do estado da arte da área de estilização artística de imagens e vídeos, realizado por Kyprianidis *et al.* em 2012 [KCWI12], as técnicas de estilização existentes podem ser divididas nas seguintes quatro categorias:

- *Stoke Based Rendering*, as quais criam iterativamente camadas virtuais de traços imitando um pincel, cujas cores, orientação e escala são derivadas de processos automáticos ou semi-automáticos de processamento de imagem;
- *Region Based Techniques*, que usam a segmentação da imagem em diversas regiões para a sua posterior estilização;
- *Renderização Baseada em Exemplo*, que tentam imitar estilos artísticos específicos através de heurísticas apreendidas para a renderização automática de novas imagens nesses mesmos estilos;
- *Filtros de Abstração de Imagens*, os quais usam técnicas de processamento de imagem para, através de operações de filtragem locais, efetuarem a renderização artística e a abstração do conteúdo visual.

Apesar de os filtros de abstração de imagens possuírem uma menor diversidade numa perspectiva artística de renderização, a utilização de uma abordagem local por parte destes filtros torna-os vantajosos, quando comparada com as restantes técnicas de renderização artística de imagens. Esta abordagem permite uma fácil adaptação dos processos de abstração de imagens a tecnologias de processamento paralelo, presentes nos CPUs modernos, assim como uma implementação direta do processo de abstração de alguns dos filtros pelos próprios GPUs. Nesse sentido, e devido ao grande número de imagens utilizadas no processo de reconhecimento facial automático a utilização de filtros de abstração de imagens torna-se mais vantajosa no âmbito desta dissertação.

3.2 Filtros de Abstração de Imagens

Os filtros de abstração de imagens permitem simplificar o conteúdo visual, destacando apenas a informação essencial contida nas imagens. Existem vários tipos de filtros de abstração, de seguida encontram-se descritas seis categorias de filtros de abstração de imagem tipicamente utilizados.

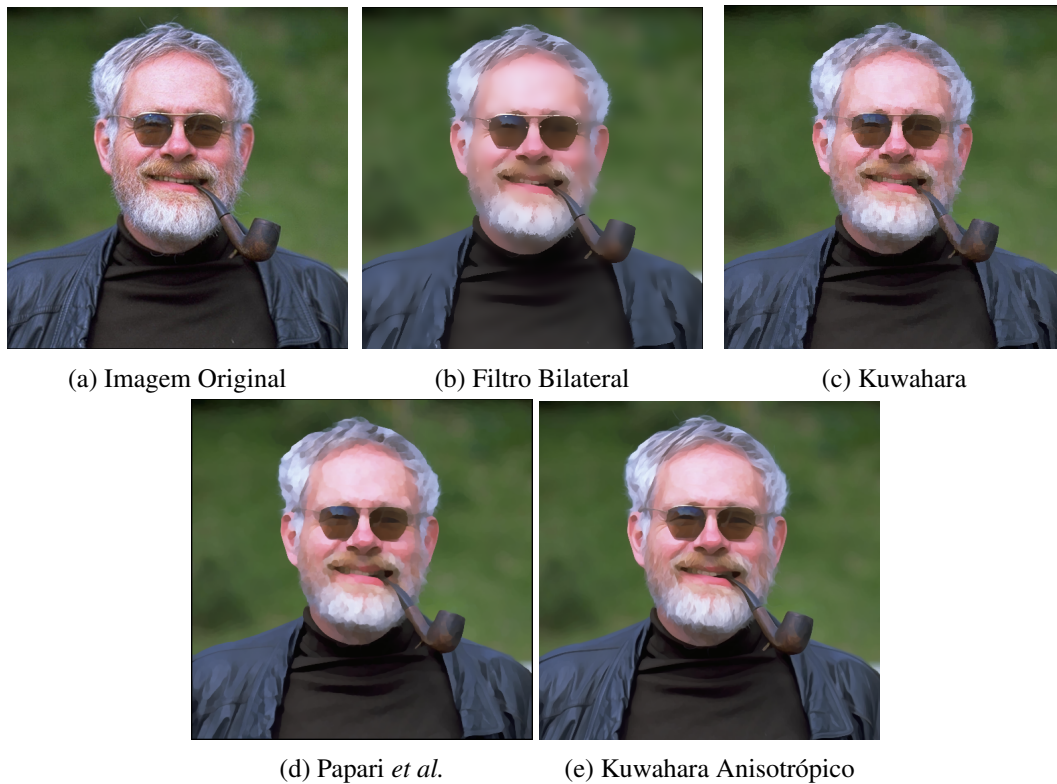


Figura 3.2: Diferentes filtros de abstração aplicados sobre a mesma imagem [KKD09].

3.2.1 Filtro Gaussiano

A utilização de filtros gaussianos no âmbito do processamento de imagem e visão por computador tem sido alvo de estudo desde há mais de 20 anos [DC93].

O principal objetivo deste filtro é suavizar a imagem original de forma a reduzir o ruído existente na mesma. Este filtro utiliza uma função gaussiana, para determinar a transformação a aplicar a cada pixel da imagem original, a qual pode ser resumida, de uma forma simplificada, como sendo uma média pesada dos valores existentes nos pixels vizinhos, na qual o peso de um pixel na média diminui conforme a sua distância ao centro aumenta.

Apesar de diminuir o ruído existente numa imagem, a filtragem gaussiana apresenta como principal desvantagem o facto de reduzir o detalhe existente na imagem, assim como os limites dos objetos nela contidos. Por esse motivo este filtro é muitas vezes utilizado paralelamente com outras técnicas de abstração visual, de modo a melhorar os resultados obtidos.

3.2.2 Filtro Bilateral

Uma abordagem baseada num filtro bilateral foi proposta em primeira mão por Tomasi e Manduchi em 1998 [TM98] e encontra-se representada na imagem 3.2b. Já em 2006, Winnemöller propôs uma melhoria dos resultados obtidos através da associação de um filtro de diferença de gaussianas na aplicação do filtro bilateral [WOG06]. Trabalhos posteriores com vista a melhorar,

os contornos produzidos pelas imagens abstraídas foram também realizados com sucesso através da introdução de um filtro de diferença de gaussianas baseado baseado em fluxo [KLC09].

O filtro bilateral caracteriza-se por suavizar as zonas com menor contraste da imagem enquanto preserva os limites de maior contraste. No entanto, os efeitos deste filtro são poucos visíveis quando toda a imagem possui um elevado contraste. Por outro lado, em imagens pouco contrastadas a informação visual removida é por vezes demasiada, causando um efeito de imagem desfocada na imagem abstraída.

3.2.3 Kuwahara

O filtro Kuwahara (fig. 3.2c) e as suas variantes, representam uma classe de filtros onde se tem verificado uma pesquisa ativa ao longo das últimas décadas [KKD09].

A ideia geral do filtro do tipo Kuwahara consiste em efetuar um achatamento da imagem, através da divisão da área a filtrar em quatro sub-regiões retangulares de menor dimensão e a posterior substituição da informação contida numa dessas regiões pela média da sub-região com menor variância. Desta forma, é removido o detalhe em zonas com elevado contraste mas são preservadas as limites das formas presentes na imagem. No entanto, na presença de ruído este filtro revela-se instável, criando artefactos na imagem. Para além disso, esta abordagem revela-se também agressiva na presença de anisotropia ¹, ou seja quando a informação direcional presente na imagem é forte (por exemplo: pelo, cabelos), não preservando essa informação na imagem abstraída.

Existem várias variantes ao filtro Kuwahara, sendo que a grande diferença entre a maior parte das variantes consiste na forma como são definidas as sub-regiões a filtrar. Papari *et al.* [PPC07], apresentaram, com bons resultados, uma variante deste filtro em que a divisão das sub-regiões é efetuada em regiões circulares, que por sua vez se encontram divididas em oito regiões de igual tamanho. Apesar das melhorias registadas na abordagem de Papari *et al.*, visíveis na figura 3.2d, esta abordagem continua a não preservar a informação direcional contida nas imagens resultando assim no aparecimento de alguns artefactos, ainda que mais pequenos e com maior definição do que na solução proposta pelo o filtro original.

Para ultrapassar as limitações do filtro Kuwahara original na presença de anisotropia, foi proposta por Kyprianidis *et al.* [KKD09], uma nova abordagem com o nome filtro Kuwahara Anisotrópico (3.2e). Nesta solução, é feita uma adaptação da forma, orientação e escala do método de seleção das sub-regiões à estrutura local da imagem, resultando numa forma que varia entre um círculo e um elipse alongada conforme a própria estrutura local da imagem. Assim, para além de preservada, a informação direcional é ainda enfatizada, resultando em imagens tipo pintura, com limites bem definidos e sem grandes artefactos visíveis.

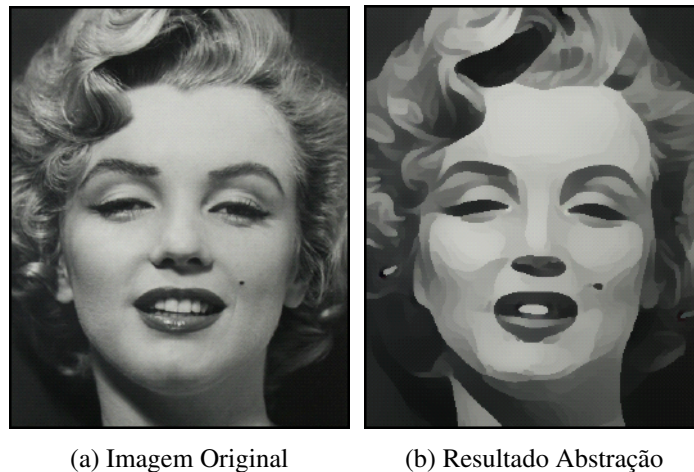


Figura 3.3: *Diffusion Shock Filter* [KL08].

3.2.4 *Diffusion Shock Filter*

Kang e Lee [KL08], foram os primeiros a utilizar *shock filtering* para a abstração de imagens através do uso de *Mean Curvature Flow (MCF)*, resultando em imagens onde para além de suavizadas as variações irrelevantes das cores, são simplificados os limites das formas presentes nas mesmas, como é possível visualizar na figura 3.3. O resultando da aplicação de *shock filtered MFC*, é, no entanto, tipicamente é muito agressivo, resultando em imagens desfocadas, onde apenas é parcialmente preservada a informação direcional contida na imagem. Esta técnica revela-se também, pouco indicada para a abstração de vídeo, uma vez que é pouco sensível a pequenas alterações nas imagens.

3.2.5 Filtros Morfológicos

Os filtros desta categoria, utilizam os princípios da morfologia matemática para reduzir o detalhe existente nas zonas coloridas, criando um efeito de uniformização da cor. A morfologia matemática foi introduzida originalmente por Matheron and Serra em 1982 [Ser82], e consiste numa teoria para a análise e processamento de estruturas geométricas. De uma forma geral maioria das operações morfológicas são baseadas em simples operações de expansão e encolhimento e são designadas dilatação e erosão.

Existe uma grande variedade de abordagens e filtros baseados nos princípios da morfologia matemática, existindo consequentemente uma grande variedade ao nível dos resultados obtidos, conforme o filtro morfológico aplicado.

3.2.6 Técnicas *Gradient Domain*

O gradiente de uma imagem representa de forma direcional as mudanças de intensidade de uma imagem independente das suas cores originais, permitindo assim uma maior flexibilidade na

¹ Anisotropia - qualidade de certos materiais cujas propriedades são diferentes consoante as direções.

manipulação das imagens. As técnicas baseadas no uso de gradientes, consistem na construção de um campo de gradiente que representa a imagem, o qual é posteriormente utilizado na reconstrução da imagem filtrada.

A grande desvantagem das técnicas baseadas no uso de gradientes é que os resultados obtidos em zonas com contraste são por vezes problemáticos e requerem correção. Para além disso, a computação deste tipo de filtragem é computacionalmente pesada e por isso não aplicável para situações de utilização em tempo real.

3.3 Trabalho Relacionado

Os filtros de abstração são normalmente utilizados por artistas para comunicar a mensagem visual mais eficazmente. No âmbito desta dissertação pretendemos analisar o impacto destes filtros não ao nível da comunicação visual efetuada, mas ao nível do impacto que estes podem causar no processo de reconhecimento facial. Esta investigação é motivada por resultados anteriores obtidos com o uso dos mesmos filtros na ilustração automática de texto.

A ilustração de texto é uma tarefa enquadrada na área da recuperação de informação multimédia e que consiste na pesquisa de imagens adequadas para a ilustração de fragmentos de texto, tais como notícias ou livros. Coelho e Ribeiro [CR12], propõem o uso da abstração de imagens para a ilustração automática de texto, através de um sistema de recuperação de informação, o qual dada uma entrada textual inicial, seleciona automaticamente uma pequena lista de imagens relacionadas a partir da análise coleção de imagens previamente processada.

O sistema de ilustração automática de texto utiliza a abstração de imagens para melhorar a informação capturada pelos descritores de informação visual utilizados e consequentemente melhorar os resultados obtidos a partir da análise baseada em conteúdos efetuada à coleção de dados.

Os resultados obtidos demonstraram que o uso de abstração tem a potencialidade de melhorar a informação recuperada, assim como reduzir as necessidades de processamento e armazenamento da coleção de dados utilizada.

Capítulo 4

Sistema de Reconhecimento Facial Visage

Intro ...

4.1 Filtros de Abstração de Imagens no Reconhecimento Facial

O reconhecimento facial em imagens sofreu uma evolução notável nos últimos 20 anos, tal como documentado no capítulo 2 deste relatório.

Em cenários cooperativos com condições de captura de imagens controladas, nomeadamente ao nível da pose, iluminação e expressões faciais, considera-se mesmo que o problema de verificação 1:1 se encontra praticamente resolvido, uma vez que as taxas de reconhecimento atingidas são satisfatórias para a grande maioria das aplicações [LJ11]. Existem também várias aplicações em situações reais com um bom nível de satisfação por parte dos seus utilizadores, como é o caso do sistema de fronteira automático dos aeroportos portugueses (ver 2.3.2), ou o controlo de entradas nas cerimónias inaugurais dos jogos olímpicos de Pequim (ver 2.3.5). Em condições específicas e favoráveis, é então possível considerar que os sistemas de reconhecimento facial automático atuais conseguem mesmo ultrapassar a capacidade de reconhecimento humana, uma vez que conseguem identificar com precisão um maior número de faces do que aquelas que um humano consegue.

Contudo, o problema de reconhecimento facial automático ainda se encontra longe de ser um problema totalmente resolvido. Em cenários onde é registada uma grande variação ao nível da pose, iluminação ou outros fatores identificados na secção 2.2.2 deste relatório, a identificação das entidades capturadas é ainda uma tarefa desafiante [LJ11]. Para além disso, a performance e satisfação obtida por parte dos utilizadores dos sistemas atuais demonstra uma grande variação tendo em conta as situações onde estes sistemas são utilizados.

Por outro lado, a crescente ubiquidade tecnológica e poder computacional presente nos diversos dispositivos utilizados no nosso dia a dia, aumenta o leque de aplicações possíveis do reconhecimento facial automático, apresentando novos desafios às soluções de atualmente existentes.(ver 2.3)

Considerando todos os fatores mencionados anteriormente e ainda o elevado valor comercial das soluções existentes e consequente falta soluções abertas, torna-se pertinente a realização de investigação na área do reconhecimento facial em imagens.

Ao nível da abstração de imagens, estudos efetuados demonstraram que aplicação destes filtros na recuperação de informação multimédia, nomeadamente no âmbito da ilustração automática de texto têm a potencialidade de melhorar a informação retornada, assim como reduzir significativamente as necessidades de processamento e armazenamento das imagens [CR12].

Tendo em conta a importância e a necessidade de investigação na área do reconhecimento facial automático, e uma vez que não existem estudos relativos à utilização de filtros de abstração no processo de reconhecimento facial, torna-se pertinente o estudo do seu impacto, no âmbito desta dissertação. Desta forma, a hipótese levantada no âmbito desta dissertação é então que o uso da abstração em imagens que vão ser alvo de reconhecimento facial pode melhorar o processo de reconhecimento facial automático em imagens.

4.2 Visage - Visão geral

A criação do sistema Visage teve como principal objetivo o desenvolvimento de um sistema de reconhecimento facial automático que permita analisar qual o impacto da abstração de imagens e outras tarefas de pré-processamento no reconhecimento facial automático. Por outro lado, com o sistema desenvolvido pretende-se ainda a criação de uma base sólida que permita o desenvolvimento de futuras aplicações que tirem partido do reconhecimento facial automático no seu funcionamento, como por exemplo, aplicações na área de *image retrieval*.

De uma forma genérica, o funcionamento do sistema Visage pode ser agrupado nas seguintes fases fundamentais:

Pré-processamento Corresponde à primeira etapa executada pelo sistema, na qual é fornecida uma galeria de imagens a pré-processar, as quais irão ser posteriormente utilizadas nas fases de treino e identificação. O pré-processamento engloba tarefas como deteção das faces na imagem, normalização do contraste e iluminação, aplicação de filtros de abstração ou alinhamento das imagens.

Treino Nesta fase, são fornecidas um conjunto de imagens anotadas ao sistema e pretende-se que este crie uma base de conhecimento que permita de futuro proceder à identificação de novas imagens de acordo com a informação treinada.

Identificação Corresponde à utilização do sistema como intuito de identificar novas imagens não anotadas. Para uma dada imagem fornecida ao sistema, este deve devolver o nome da pessoa que se encontra na imagem tendo em conta a informação obtida na fase de treino. Nesta fase é também possível obter uma lista de entidades, ordenada de forma crescente pela similaridade em relação à pessoa representada na imagem.

Uma vez que foco desta dissertação não visa o desenvolvimento de algoritmos de reconhecimento facial, mas sim a análise do impacto da manipulação efetuada sobre as imagens a reconhecer, foram utilizadas bibliotecas de código aberto que permitem agilizar o desenvolvimento do sistema. Por outro lado, a utilização destas bibliotecas permite também a contribuição para uma área onde as soluções abertas existentes são ainda poucas.

Para a manipulação das imagens e implementação do sistema de reconhecimento facial foi então utilizada a biblioteca *Open Source Computer Vision Library (OpenCV)*, a qual é uma biblioteca de código aberto nas áreas de visão por computador e *machine learning*, onde se encontram implementados de mais de 2500 algoritmos relacionados com as áreas de computação gráfica e visão por computador. Esta biblioteca possui uma comunidade de mais de 47 mil pessoas, já registou mais de 5 milhões de downloads e é utilizada globalmente por empresas como a Google, Yahoo, Microsoft, Intel, IBM, Sony, Honda, Toyota [its]. Ao nível do reconhecimento facial, esta biblioteca disponibiliza um módulo denominado *Face Recognizer*, onde se encontram implementados os algoritmos *Eigenfaces*, *Fisherfaces* e *Local Binary Patterns Histograms*. Tendo em conta a ampla utilização da biblioteca OpenCV e a sua constante atualização pela comunidade, assim como as facilidades providenciadas pelo módulo de reconhecimento facial, esta foi considerada a biblioteca ideal para utilizar como base de implementação do sistema de reconhecimento facial a criar.

Ao nível dos filtros de abstração foram utilizados os filtros gaussiano e bilateral, implementados na biblioteca *OpenCV*, e ainda o filtro anisotrópico kuwahara, recorrendo para a sua utilização à implementação efetuada por Kyprianidis *et al.* [KKD09] do mesmo filtro.

Finalmente, ao nível das galerias de imagens, foi utilizada a coleção *Labeled Faces in the Wild* descrita na secção seguinte deste relatório.

4.3 Coleção de dados - *Labeled Faces in the Wild*

A evolução do estado da arte do reconhecimento facial automático em imagens tem beneficiado em larga escala do aumento constante dos recursos disponíveis para o seu estudo, nomeadamente através da criação de novas e mais completas coleções de dados [HRBLm07]. A grande maioria destas coleções, caracteriza-se por ter condições de captura controladas com o intuito de efetuar o estudo de fatores específicos que afetam a qualidade do reconhecimento. No âmbito desta dissertação, temos em vista a análise do problema de reconhecimento facial automático do um ponto de vista geral, e próximo das situações onde este tipo de sistemas são utilizados verdadeiramente, desta forma, a escolha de uma biblioteca apropriada e que represente as grandes variações associadas aos diferentes desafios que afetam o reconhecimento facial em imagens torna-se crucial.

Tendo em conta as preocupações acima mencionadas, para o desenvolvimento do sistema de reconhecimento facial Visage e posterior análise dos resultados obtidos no âmbito nesta dissertação foi escolhida a coleção de imagens *Labeled Faces in the Wild*.

A coleção *Labeled Faces in the Wild (LFW)* é uma base de dados fotográfica desenhada especificamente para o estudo do problema de reconhecimento facial, particularmente em situações onde as condições de captura das imagens não possuem restrições. Nesse sentido, as imagens que constituem a galeria caracterizam-se por possuir uma grande variabilidade, nomeadamente ao nível da pose, expressão, iluminação, etnia, idade, género, vestuário e qualidade da câmara com a qual foi efetuada a captura[HRBLm07].

Nesta coleção encontram-se representados um conjunto de 5749 indivíduos, ilustrados por 13233 imagens. O número de imagens por pessoa apresenta no entanto uma grande variação, desde um mínimo de 1 imagem por pessoa para 4069 indivíduos, até um máximo de 530 imagens que representam o antigo presidente dos Estados Unidos da América George W. Bush. O gráfico 4.1 ilustra a grande variação da distribuição do número de imagens por pessoa. No eixo vertical encontram-se representadas o número de pessoas em escala logarítmica e no eixo horizontal o número de imagens correspondente.

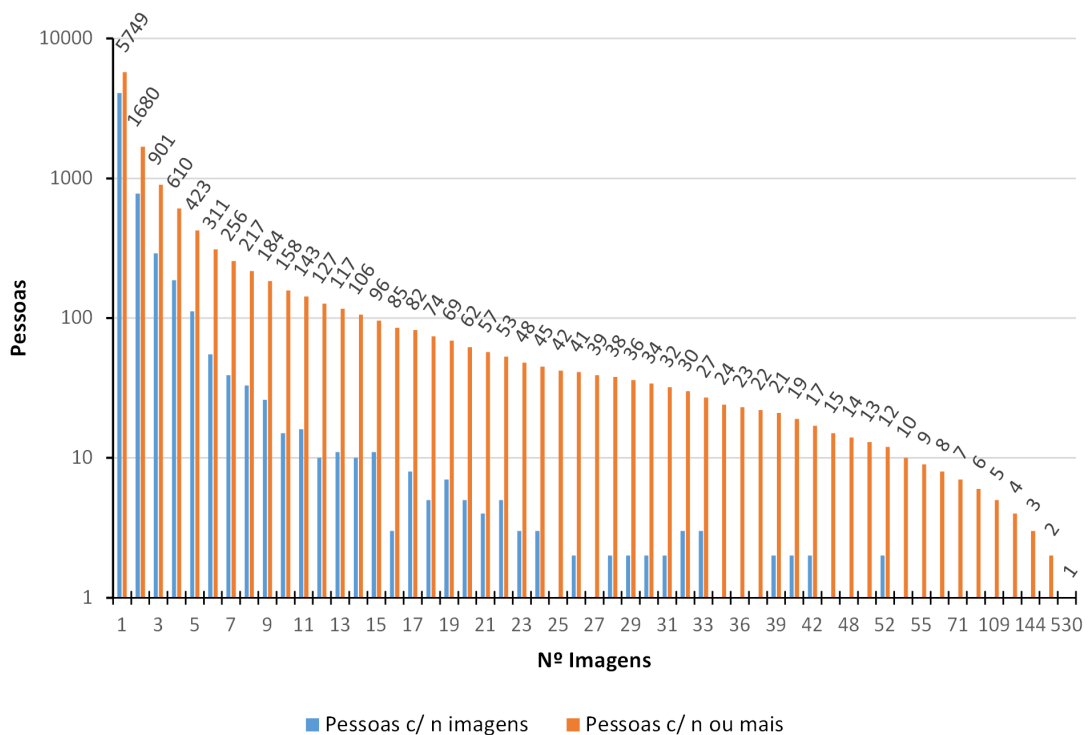


Tabela 4.1: Caracterização da Biblioteca LFW

Característica	Valor
Imagens	13233 imagens
Pessoas representadas	5749 pessoas
Tamanho total / Formato	1.18 GB / PNG
Tamanho de cada imagem: min / médio / max	46.49 KB/ 93.72 KB/ 147.89 KB
Dimensões imagem	250×250 pixels

da pessoa representada. A tabela 4.1 sintetiza algumas das características mais significativas da biblioteca.

Para além da coleção LFW original, encontram-se também disponíveis para fins de investigação duas outras coleções derivadas da coleção original, LFW funneled e LFW-a, onde as imagens originais foram sujeitas a um pré-processamento com vista a efetuar o alinhamento da posição da face na imagem, cada uma correspondendo a uma técnica diferente de alinhamento da imagem, respetivamente. Um exemplo dos resultados obtidos nas diferentes bibliotecas pode ser visto na figura 4.2.



Figura 4.2: Diferentes versões da biblioteca LFW

A coleção LFW é particularmente interessante para o estudo do desempenho do sistema desenvolvido e consequentemente do impacto do uso da abstração de imagens no reconhecimento facial uma vez que a variabilidade presente nas suas imagens, representa uma amostra fiável das variações encontradas no dia-a-dia de uma pessoa. Por outro lado, o facto de as suas imagens não terem sido capturadas com o intuito específico da análise do problema do reconhecimento facial, aproxima este conjunto de dados das imagens utilizadas por um utilizador final de um sistema de reconhecimento facial automático. Finalmente, e uma vez que a tarefa de alinhamento da face na imagem não é um objeto de estudo primordial desta dissertação, a existência de versões já alinhadas da coleção permitiu agilizar os testes realizados, retirando um fator importante de variação dos resultados obtidos. Assim sendo, no âmbito desta dissertação optamos pela utilização da versão alinhada LFW-a para fins de avaliação do desempenho do sistema de reconhecimento facial criado e do impacto do uso de filtros de abstração sobre essas imagens.

4.4 Arquitetura

O sistema de reconhecimento facial criado é constituído pelos seguintes 6 módulos:

Person Módulo responsável por efetuar o encapsulamento da informação relativa a uma pessoa, nomeadamente o seu nome, código de identificação e imagens a esta associada. De modo a facilitar a avaliação do sistema criado é ainda possível armazenar de forma diferenciada imagens de treino e teste de cada pessoa.

Library Módulo responsável por encapsular a informação de uma galeria de imagens utilizada para treinar o sistema, nomeadamente o seu diretório, as pessoas representadas nessa biblioteca (agrupadas num *array* de objetos do tipo *Person*) e alguns dados estatísticos como o número mínimo e máximo de imagens existentes por pessoa e o número total de imagens da galeria. Apesar de poder ser utilizado com diferentes funcionalidades este módulo foi desenvolvido com o intuito de facilitar a avaliação do sistema Visage com diferentes galerias de imagens, pelo que na criação de uma *Library* é possível definir qual a percentagem das imagens, de cada pessoa, que devem ser utilizadas como conjunto de teste e conjunto de treino do sistema (Ver capítulo 5 para mais detalhes).

FaceDetector A par do módulo *FaceModel* é um dos componentes principais do sistema, sendo responsável por efetuar um conjunto de tarefas de pré-processamento sobre as imagens utilizadas. Este módulo encontra-se descrito detalhadamente na secção 4.5.

FaceModel O módulo *FaceModel* constitui uma camada de nível superior implementada para encapsular a comunicação com módulo de reconhecimento facial disponível na biblioteca *OpenCV*, designado *FaceRecognizer*, assim como estender algumas das funcionalidades do mesmo. Este módulo encontra-se descrito com maior detalhe na secção 4.6.

Evaluator Módulo responsável por agrupar os métodos e objetos utilizados para efetuar a avaliação do sistema.

Utils Engloba algumas funções úteis utilizadas para o desenvolvimento e avaliação do sistema Visage que não se enquadram no âmbito dos restantes módulos.

4.5 Módulo *FaceDetector*

O módulo *FaceDetector* é um dos componentes principais do sistema de reconhecimento facial criado, sendo responsável por efetuar um conjunto de tarefas de pré-processamento sobre as imagens fornecidas ao sistema.

Na figura 4.3 encontram-se representadas as diferentes tarefas de pré-processamento disponíveis no sistema de reconhecimento facial Visage, assim como a ordem pela qual as mesmas são aplicadas. As etapas de normalização, aplicação do filtro de abstração e aplicação da máscara

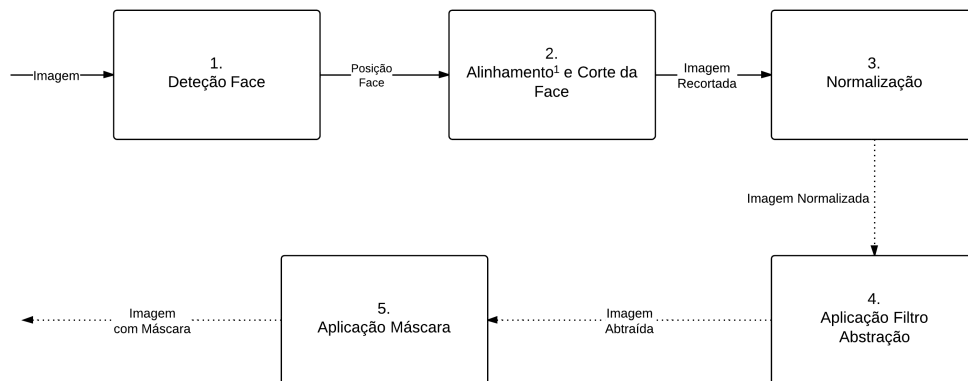


Figura 4.3: Diferentes tarefas de pré-processamento disponíveis no sistema Visage.

constituem tarefas de pré-processamento opcionais pelo que se encontram representadas por uma linha a tracejado. ¹

4.5.1 Detecção Face

A existência de elementos de fundo numa imagem produz um impacto significativo nos resultados obtidos por sistemas de reconhecimento facial automático, pelo que a deteção da zona de uma imagem onde se encontra representada a face a identificar é fundamental para um bom desempenho do sistema criado. A fase de pré-processamento designada de "deteção face" implementada no sistema Visage tem então como objetivo, para uma dada imagem fornecida ao sistema, determinar qual a zona da imagem onde se encontra a face a identificar.

A deteção facial é efetuada com recurso um classificador em cascata, designado de *boosted cascade classifier*, treinado especificamente para a deteção de faces. O classificador utilizado encontra-se implementado na biblioteca *OpenCV* e baseia-se nos trabalhos de Viola e Jones [VJ01] e posteriores melhorias introduzidas por Lienhart, Kuranov e Pisarevsky [LKP03].

¹Tal como mencionado em 4.3, na avaliação do sistema foi utilizada a versão alinhada da biblioteca LFW, designada de LFW-a, pelo que a tarefa de alinhamento das imagens não foi realizada na avaliação de desempenho reportada nesta dissertação.

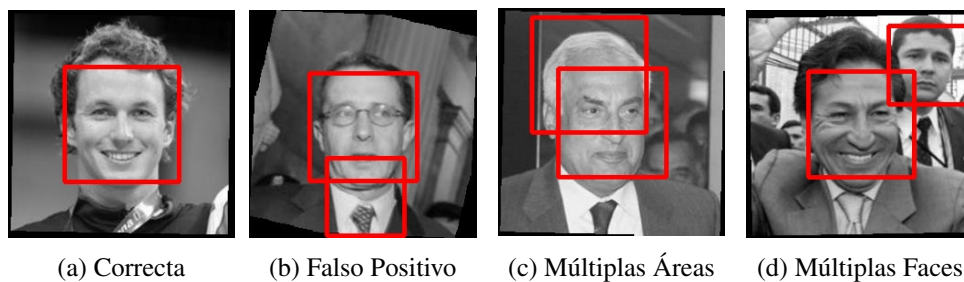


Figura 4.4: Comparação entre face correctamente detectada e situações de erro na deteção

Um classificador deste tipo é treinado com algumas centenas de imagens, de um objeto particular, neste caso uma face em posição frontal, designando-se essas imagens de exemplos positivos; assim como algumas imagens arbitrárias e não relacionadas, designadas de exemplos negativos. Quer os exemplos negativos, quer os positivos são previamente escalados para o mesmo tamanho. Após o treino, o classificador é capaz de determinar se numa região de igual tamanho às imagens utilizadas no seu treino, designada de janela, é provável, ou não, a existência de uma face. Para efetuar a pesquisa de faces em toda a imagem a janela do classificador é movida ao longo da imagem de forma a que toda a área existente seja pesquisada. Para a pesquisa de faces de diferentes tamanhos é possível escalar o classificador, efetuando o varrimento da imagem nas diversas escalas que se pretende pesquisar.

Um dos principais contributos de Viola e Jones para a construção de um sistema de deteção facial eficaz prende-se com utilização de um algoritmo de aprendizagem, designado de *Adaboost*, que efetua a seleção de um número de características visuais críticas comuns a um conjunto alargado de imagens, permitindo assim a extração de pormenores essenciais para a classificação de um tipo de objecto [VJ01]. Desta forma é possível a criação de vários classificadores simples focados em características específicas da imagem. A palavra "*boosted*" no classificador utilizado deriva da utilização desta técnica no mesmo.

A designação de classificador em cascata deriva da utilização de um conjunto de classificadores simples, de forma sequencial e combinada. Para a determinação da existência de uma face, todos os classificadores utilizados têm de aprovar a existência da mesma, caso contrário a face é rejeitada. Um exemplo básico de um classificador seria uma árvore de decisão com pelo menos duas folhas.

No sistema Visage é utilizado um classificador facial, disponibilizado pela biblioteca *OpenCV*, treinado através das características obtidas pela extração das *Local Binary Patterns* [AHP06] de um conjunto de imagens de caras frontais. Contudo, a implementação existente na biblioteca *OpenCV* deste tipo de classificadores pode ser igualmente utilizada para reconhecer qualquer tipo de objetos para os quais exista informação de treino disponível.

Remoção Conflitos e Falsos Positivos

Para uma dada imagem, a tarefa de deteção face é responsável pela determinação de uma área de interesse, de forma rectangular, onde se encontra a face que se pretende fornecer às etapas

Tabela 4.2: Resultados etapa de de detecção facial para todas as pessoas com pelo menos 2 imagens na biblioteca LFW-a.

Pessoas	1680
Imagens	9165
Imagens com faces detectadas s/ conflito	95.50% (8753 imagens)
Conflitos e falsos positivos resolvidos automaticamente	1.63% (149 imagens)
Conflitos e falsos positivos não resolvidos (empates)	0.34% (31 imagens)
Imagens com faces não detectadas	2.52% (231 imagens)

de treino ou teste do sistema Visage. Um exemplo de uma imagem cuja face foi corretamente detetada e da respetiva área de interesse, assinada a vermelho, pode ser encontrado na figura 4.4a.

Apesar do bom desempenho do classificador utilizado na detecção das faces presentes nas imagens existem algumas situações onde se verificou a existência de erros ou conflitos nos resultados obtidos, tal como ilustrado em 4.4. As três situações de erro encontradas podem ser agrupadas em:

Falsos Positivos Zonas da imagem que não contém faces podem ser assinaladas como área de interesse, tal como ilustrado em 4.4b;

Múltiplas Áreas de Interesse Para a mesma face, na análise de uma imagem são devolvidas várias regiões de interesse, tal como pode ser visto em 4.4c;

Múltiplas Faces Na galeria utilizada para o desenvolvimento e avaliação do sistema, LFW-a, existem fotografias onde mais do que uma pessoa se encontra representada na mesma imagem, no entanto, em cada imagem é apresentada apenas a anotação textual para a face mais relevante na fotografia, tal como pode ser visto em 4.4d, onde apenas a face maior possui uma anotação textual associada, pelo que apenas essa face deve ser extraída da imagem.

Para a remoção dos casos de conflito observados na detecção das faces presentes na biblioteca LFW-a e ilustrados em 4.4, foi então adoptada a seguinte estratégia:

Para uma dada imagem é verificada a existência de uma cara usando o classificador em cascata treinado para o reconhecimento de caras frontais. Caso se verifique a detecção de um conflito, nessa imagem é necessário proceder à re-avaliação de cada uma das regiões assinaladas como área de interesse, de forma a determinar em qual se encontra a face que se pretender utilizar.

Na avaliação das várias regiões de interesse é procurada a existência de olhos, boca e nariz nas mesmas, com recurso a três classificadores em cascata treinados para o reconhecimento de cada um destes componentes em específico. Posteriormente é calculada uma pontuação para cada uma das regiões de interesse conforme os elementos encontrados:

- +100 Pontos caso sejam encontrados o par de olhos (olho esquerdo e direito) na face;
- +50 Pontos caso seja encontrada a boca. No caso de terem sido encontrados previamente os olhos, é também verificado se distância entre a boca e os olhos é de pelo menos 20 pixels, de modo a garantir que um olho não é classificado como boca e vice-versa;

- + 50 Pontos caso seja encontrado o nariz na face.

A pontuação de uma região de interesse varia assim entre os 0 e os 200 pontos. No final, é guardada apenas a região de interesse com maior pontuação. Em caso de empate, são guardadas ambas as regiões de interesse, e o utilizador do sistema é avisado do conflito, o qual deverá ser resolvido posteriormente de forma manual.

O desempenho da tarefa de detecção facial numa amostra da biblioteca LFW-a, pode ser vista na tabela 4.2. Nesta amostra foram incluídas apenas as fotografias de pessoas que tenham pelos menos 2 imagens na biblioteca LFW-a, uma vez que apenas essas poderão ser utilizadas na avaliação do desempenho do sistema, utilizando, pelo menos, uma imagem para treino e uma imagem para teste. Tal como é possível visualizar o desempenho desta etapa é muito satisfatório, sendo que 95.5% das faces imagens presentes nas imagens são detetadas automaticamente, às quais acrescentem 149 imagens para as quais foi possível resolver automaticamente os conflitos detetados.

4.5.2 Alinhamento e Corte

A tarefa de detecção face é responsável pela determinação de uma região de interesse na imagem original onde se encontra a face, sendo essa região correspondente a um retângulo tal como ilustrado na figura 4.4a. Após a detecção da região de interesse onde se encontra a face, é então possível proceder ao seu alinhamento e corte através da seguinte estratégia:

1. Utilizando um classificador em cascata existente na biblioteca *OpenCV* é verificada a posição do olho esquerdo na região de interesse determinada para a face;
2. Utilizando um classificador em cascata existente na biblioteca *OpenCV* é verificada a posição do olho direito na região de interesse determinada para a face;
3. São removidos conflitos e falsos positivos nas posições obtidas para o olho esquerdo e direito (ver sub-secção seguinte), obtendo-se assim as duas regiões mais prováveis para cada um dos olhos;
4. É calculado o ponto central das regiões determinadas como mais prováveis para o olho esquerdo e direito, em que $p_{esquerdo}(x_1, y_1)$ corresponde ao ponto central da região detetada para o olho esquerdo, e $p_{direito}(x_2, y_2)$ corresponde ao ponto central da região detetada para o olho direito.
5. É calculado o ângulo de rotação da imagem (θ), a partir da posição determinada para os olhos:

$$\theta = \frac{\arctan(x_1 - x_2, y_1 - y_2) * 180.0}{\Pi} \quad (4.1)$$

6. A imagem é rodada de acordo com o ângulo de rotação θ calculado.
7. A imagem é cortada de acordo com a região de interesse determinada na tarefa de detecção face.

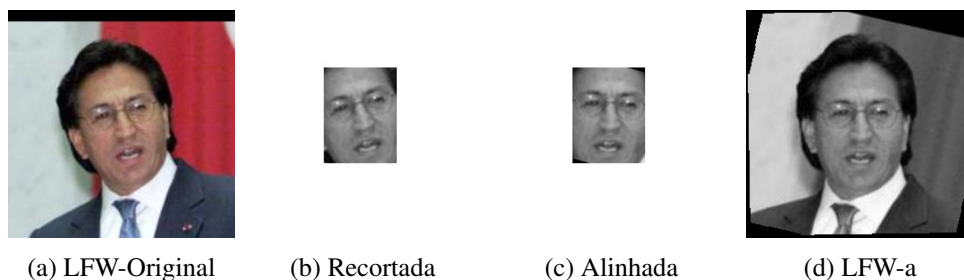


Figura 4.5: Comparação entre versões original, recortada, alinhada e LFW-a.

Caso não se pretenda efetuar o alinhamento da face, mas apenas a sua segmentação da restante imagem, como se verifica na utilização da galeria já alinhada, LFW-a, na avaliação do desempenho do sistema, os passos 1 a 6 são ignorados, sendo apenas segmentada a imagem de acordo com a região de interesse determinada na tarefa de deteção face. Na figura 4.5 encontra-se representada uma comparação entre uma imagem apenas recortada, a mesma imagem alinhada e recortada pelas estratégias acima descritas e ainda as versões LFW e LFW-a dessa imagem.

Remoção Conflitos e Falsos Positivos

Tal como explicado anteriormente, a necessidade de alinhamento de uma face previamente detetada é calculada com base na localização estimada para o seu olho esquerdo e direito e a inclinação da reta formada pela intersecção dos dois pontos centrais de cada um dos olhos. Para detetar um olho, à semelhança da estratégia adotada para detetar uma face, é utilizado um classificador em cascata que devolve uma região rectangular onde é provável a existência desse mesmo olho.

A morfologia de um olho é um caso particularmente difícil de deteção, tendo-se registado a existência de um grande número de falsos positivos para os classificadores utilizados (olho esquerdo e direito), assim como uma má diferenciação entre o olho esquerdo, direito e a boca em algumas das imagens, pelo que a adoção de uma estratégia de remoção de conflitos se revelou essencial para a conclusão desta tarefa de pré-processamento com sucesso. A estratégia adoptada foi então a seguinte:

Considerando w a largura de uma imagem de face a alinhar e h a sua altura. Sejam $c_{esquerdo}(x_1, y_1)$ e $c_{direito}(x_2, y_2)$ os pontos correspondentes aos cantos superiores esquerdos das regiões de interesse detetadas para o olho esquerdo e direito, respetivamente, e $w_{esquerdo}$, $w_{direito}$, $h_{esquerdo}$ e $h_{direito}$ a largura e altura dessas mesmas regiões. Os conflitos entre as diversas regiões de interesse devolvidas pelos classificadores utilizados são removidos através da utilização das seguintes restrições:

- $x_1 < \frac{w}{4}$ e $\frac{w}{4} < x_2 < \frac{3 \times w}{4}$;
- $y_1 < \frac{h}{2}$ e $y_2 < \frac{h}{2}$;
- $w_{esquerdo} < \frac{5 \times w}{8}$ e $w_{direito} < \frac{5 \times w}{8}$;
- $h_{esquerdo} < \frac{h}{2}$ e $h_{direito} < \frac{h}{2}$;

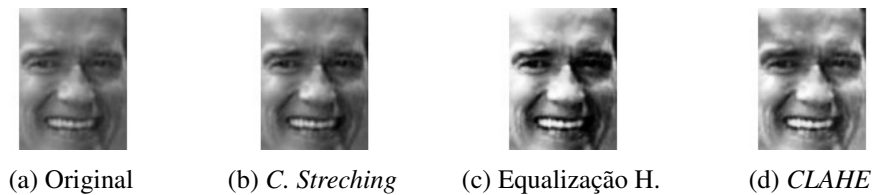


Figura 4.6: Comparação entre imagem original e as várias técnicas de normalização disponíveis.

4.5.3 Normalização Contraste

Tal como destacado na secção 2.2.1 a tarefa de normalização encontra-se presente em alguns sistemas de reconhecimento facial automático com o objetivo de melhorar os resultados obtidos independentemente de fatores como a iluminação ou pose das faces nas imagens. A tarefa de normalização do contraste de imagem tem como objetivo normalizar a face no que diz respeito à sua iluminação e aumentar o contraste existente na mesma. De forma a tornar o sistema versátil e analisar o impacto de diferentes estratégias foram implementadas as seguintes alternativas, ilustradas na figura 4.6, para a normalização do contraste:

Contrast Stretching A técnica de *Contrast Stretching*, também designada por alguns autores simplesmente de normalização, consiste numa tentativa de alargar a gama de valores utilizados para codificar a intensidade de uma imagem de forma a que seja utilizada toda a gama de valores possíveis. No sistema desenvolvido as imagens utilizadas encontram-se convertidas para tons de cinzento, pelo que a normalização efetuada consiste num mapeamento linear dos valores originais para uma gama de intensidade entre 0 e 255. Um exemplo de uma face normalizada segundo esta estratégia pode ser visualizado na imagem 4.6b.

Equalização Histograma O histograma de uma imagem descreve a distribuição estatística dos níveis de intensidade de uma imagem, neste caso concreto dos níveis de cinzento, em função do seu número de pixels. A equalização do histograma de uma imagem consiste no mapeamento das variações da escala de cinzentos de forma a que o histograma resultante se aproxime de uma outra distribuição, mais alargada e idealmente uniforme na distribuição dos valores de intensidade da imagem. Este procedimento parte do princípio de que a qualidade da imagem é uniforme em toda a imagem, aplicando um mapeamento similar a toda a imagem [BK08]. Um exemplo de uma imagem cujo histograma foi equalizado pode ser visto na figura 4.6c.

CLAHE *Contrast limited adaptive histogram equalization (CLAHE)* procura ultrapassar as limitações da equalização de contraste, através de uma abordagem local ao problema de normalização do histograma de uma imagem. Ao contrário da abordagem tradicional apresentada acima, esta técnica tem em conta a intensidade de um conjunto de pixels e dos seus vizinhos para a normalização do seu contraste e não a intensidade de toda a imagem. Para além disso, tal como o seu nome indica, esta abordagem inclui ainda um limite para o qual a equalização do histograma é realizada, evitando assim o mapeamento demasiado agressivo

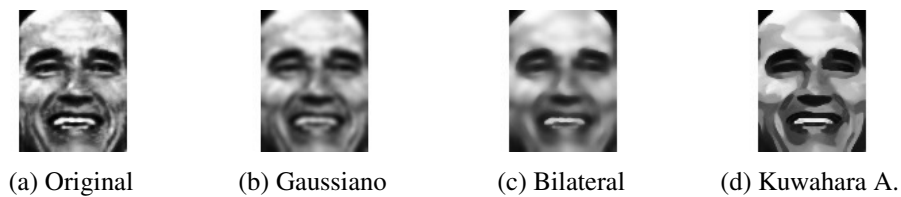


Figura 4.7: Comparação entre imagem original e os vários filtros de abstração utilizados.

quando duas escalas de intensidade dispares se encontram na mesma região [Rez04]. Na figura 4.6d encontra-se representada uma imagem resultante da sua normalização através da técnica CLAHE.

4.5.4 Aplicação Filtro Abstração

A quarta etapa de pré-processamento efetuada no sistema Visage corresponde à aplicação de um filtro de abstração sobre uma imagem. Os filtros de abstração constituem uma forma moderna de abstração de conteúdo visual, com a sua utilização no âmbito do sistema Visage, pretendemos analisar se o seu uso potencia o desempenho do sistema ao simplificar o conteúdo visual, destacando apenas a informação essencial presente numa imagem.

Nesta etapa encontram-se disponíveis três filtros, Gaussiano, Bilateral e Kuwahara Anisotrópico, os quais constituem três níveis de abstração sucessivamente mais complexos. A abstração efetuada para cada filtro encontra ilustrada na figura 4.7, apresentado-se de seguida uma breve descrição de cada um dos filtros utilizados:

Filtro Gaussiano A aplicação de um filtro gaussiano para a suavização de uma imagem é uma técnica amplamente utilizada em áreas próximas do processamento de imagem e computação gráfica e tem como objetivo de reduzir o ruído e detalhe de uma imagem. Um filtro gaussiano calcula uma média pesada dos valores dos pixels vizinhos, na qual o peso de um pixel na média diminui conforme a sua distância ao centro aumenta.

No âmbito do sistema Visage a utilização deste filtro tirou partido da sua implementação existente na biblioteca *OpenCV*, nomeadamente a partir da função *gaussianBlur*, tendo sido utilizada uma vizinhança de raio sete pixels.

Filtro Bilateral Tal como introduzido na revisão do estado da arte efetuada em 3.2.2, o filtro bilateral tem como objetivo suavizar as zonas com menor contraste na imagem sem afetar os limites de maior contraste, permitindo assim reduzir o ruído presente em uma imagem enquanto os seus contornos são preservados.

À semelhança do filtro gaussiano a utilização deste filtro foi efetuada com recurso à implementação existente na biblioteca *OpenCV*, nomeadamente a partir da função *bilateralFilter*, tendo sido utilizado um raio de oito pixels vizinhos na aplicação do filtro.

Filtro Kuwahara Anisotrópico O filtro Kuwahara Anisotrópico (FKA) consiste numa generalização do filtro Kuwahara (ver 3.2.3) que remove alguns artefactos originados na aplicação

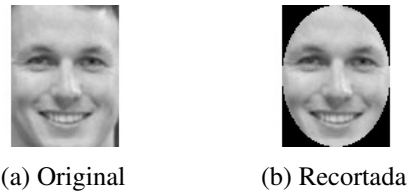


Figura 4.8: Exemplo de face já segmentada, com e sem máscara.

do filtro original através da adaptação da forma, escala e orientação do filtro à estrutura local das características da imagem [KKD09]. Desta forma é produzido um efeito de abstração tipo pintura, onde é removida informação não essencial em zonas de elevado contraste, enquanto são preservados os limites representados nas zonas de menor contraste. As imagens ficam assim com a clareza de uma ilustração, mas preservam a informação direcional tal como nas pinturas a óleo clássicas. Por outro lado, este filtro tira partido da placa gráfica para a realização da abstração das imagens, tornando-se assim particularmente indicado para o processamento de um elevado número de fotografias.

Para a utilização deste filtro foi utilizada a implementação de Jan Kyprianidis disponível em [] ².

4.5.5 Aplicação Máscara

A existência de fundo e elementos externos à face numa imagem pode afetar significativamente os resultados obtidos no seu reconhecimento ao permitir que um conjunto de imagens com fundos semelhantes possam ser identificadas como pertencendo à mesma pessoa não pela face representada na imagem, mas pelos elementos presentes no fundo da mesma. A aplicação de uma máscara elíptica sobre face visa a remoção do fundo ainda existente na imagem após a deteção e segmentação da face e eventuais tarefas de pré-processamento realizadas.

Um exemplo de uma imagem antes e depois da aplicação da máscara pode ser visto em 4.8. A estratégia de aplicação de uma máscara para melhoria dos resultados obtidos no reconhecimento facial já foi aplicada com sucesso anteriormente em diversas situações, como são exemplo a avaliação FERET [PRR00] ou o trabalho realizado por Ahonen *et al.* [AHP04].

4.6 Módulo *FaceModel*

O módulo *FaceModel* constitui uma camada de nível superior implementada para encapsular a comunicação com módulo de reconhecimento facial disponível na biblioteca *OpenCV*, designado *FaceRecognizer*, assim como estender algumas das funcionalidades do mesmo.

²<https://code.google.com/p/gpuakf/>

As principais responsabilidades do módulo *FaceModel* consistem em criar e treinar de um modelo de reconhecimento facial, assim como determinar para uma dada imagem qual a sua identificação mais provável, ou qual o conjunto ordenado de identificações mais prováveis a partir do modelo criado.

O modelo de reconhecimento facial pode ser criado com três algoritmos de reconhecimento facial diferentes, *Eigenfaces*, *Fisherfaces* e *LBPH*, implementados na biblioteca OpenCV e descritos em mais pormenor nas secções 4.6.1, 4.6.2 e 4.6.3, respetivamente. Uma vez treinado um modelo é também possível armazenar o mesmo num ficheiro e carrega-lo posteriormente, evitando assim um novo treino, o qual dependendo do número de imagens a treinar pode revelar-se computacionalmente custoso.

As funcionalidades disponíveis na biblioteca *OpenCV* para os três algoritmos implementados foram expandidas no de forma a permitir que, para uma dada imagem fornecida ao sistema, seja devolvida uma lista ordenada, por ordem crescente de similaridade, de entidades. Desta forma é possível para além de identificar a pessoa presente numa imagem, dizer qual a lista de pessoas que mais se parecem com a pessoa representada na imagem. Uma descrição detalhada da expansão efetuada às funcionalidades disponíveis na biblioteca OpenCV pode ser encontrada em 4.6.4.

4.6.1 *Eigenfaces*

O método *Eigenfaces* foi introduzido por Turk e Pentland em 1991 [TP91], e tira partido da Análise dos Componentes Principais (ACP) para efetuar o reconhecimento facial automático.

A análise de componentes principais tem como objetivo determinar as relações existentes entre diferentes conjuntos de dados, nomeadamente ao nível das suas diferenças e semelhanças, tirando partido da redundância existente para criar uma representação reduzida dos dados sem que a perda de informação ocorrida seja significativa. As imagens faciais possuem uma grande redundância natural, o algoritmo *Eigenfaces*, através da análise dos componentes principais dessas imagens, efetua uma projeção das imagens faciais num sub-espaco onde se evidenciam apenas as variações entre as diversas caras conhecidas pelo sistema.

A redução do espaço de representação revela-se fulcral no problema de reconhecimento facial em imagens, devido à grande dimensionalidade exigida para a representação de uma face. Considerando, por exemplo, uma dada imagem de $n \times m$ pixels de tons cinzentos, essa imagem poderia ser traduzida por um espaço vetorial de $m = n \times m$ pixels, assim sendo, uma imagem de apenas 256×256 pixels necessitaria de um total de 65536 pixels para ser representada.

O processo de reconhecimento facial com recurso ao algoritmo *Eigenfaces* consiste nos seguintes passos:

1. Aquisição de um conjunto de dados iniciais (conjunto de treino);
2. Projeção dos dados obtidos num sub-espaco de faces através da ACP;
3. Aquisição de uma imagem a reconhecer;

4. Projeção da face a reconhecer no sub-espço do conjunto de treino, calculando as distâncias obtidas para cada face conhecida pelo sistema;
5. Determinar qual a face do conjunto de treino com menor distância à face a reconhecer;
6. Caso a distância para a face obtida seja menor do que um limite operacional estabelecido, a imagem é reconhecida como sendo essa pessoa, caso contrário, a face é identificada como sendo uma pessoa desconhecida pelo sistema.

Este método efetua assim uma abordagem holística ao problema de reconhecimento facial em imagens, uma vez que tem em consideração a representação facial como um todo, não fazendo a distinção entre pontos específicos da face como olhos, orelhas ou nariz para efetuar o reconhecimento facial. Uma vantagem deste tipo de representação é a reduzida sensibilidade ao ruído presente nas imagens [ZCPR03].

4.6.2 *Fisherfaces*

A análise dos componentes principais visa determinar o sub-espço onde se verifica uma maior variação entre um conjunto de imagens. No entanto, a variação na representação facial de uma pessoa encontra-se muitas vezes relacionada com mudanças de expressões faciais ou iluminação dos indivíduos, pelo que, o sub-espço criado pelo algoritmo *Eigenfaces* não traduz muitas vezes apenas as diferenças de identidade entre os diversos indivíduos, mas também as diferenças verificadas entre as várias representações de um indivíduo devido à variação nas condições de captura das imagens. O algoritmo *Fisherfaces* [BHK97, EC97, ZCK98], tenta resolver este problema, através da aplicação de um passo de Análise Linear Discriminante (ALD) (*Linear Discriminant analysis*) após a análise dos componentes principais, de forma a determinar mais corretamente as variações intra-classe existentes no conjunto de imagens a avaliar.

A ALD, inicialmente introduzida por Fisher em 1936 [Fis36], tenta maximizar as diferenças existentes entre diferentes indivíduos(inter-classe) e minimizar as variações entre imagens da mesma pessoa(intra-classe) de modo a obter uma representação mais robusta em termos de variação ao nível da iluminação. Após a aplicação dos passos de ACP e ALD, o processo de reconhecimento do método *Fisherfaces* é semelhante ao efetuado pelo método *Eigenfaces*, sendo também efetuada uma abordagem holística ao reconhecimento facial.

4.6.3 *Local Binary Patterns Histograms (LBPH)*

Ao contrário dos algoritmos descritos anteriormente, o LBPH efetua uma abordagem local ao problema de reconhecimento facial, efetuando uma extração das características locais de uma imagem. Este tipo de abordagem possui a vantagem de possuir uma baixa dimensionalidade implícita, pelo que não existe a necessidade de efetuar a projeção das imagens num sub-espço.

A ideia base das *Local Binary Patterns (LBP)*, consiste em resumir a estrutura local de imagem através de uma comparação de um pixel com os seus vizinhos. Dado um pixel central é analisada a diferença entre esse pixel e cada um dos seus vizinhos. Se a intensidade do pixel for maior ou

igual do que a do seu vizinho é atribuído o valor 1, caso contrário é atribuído o valor 0. Cada pixel pode então ser traduzido por um número binário do género 11001111. Dados 8 pixeis é então possível efetuar 2^8 combinações, cada uma designada de LBP.

Ahonen *et al.* [AHP04], propuseram o uso de LBP para o reconhecimento facial em imagens. A sua abordagem consiste na divisão da face em pequenas regiões das quais são derivadas LBP e concatenadas numa representação única designada de LBPH, a qual representa uma imagem facial. O reconhecimento é depois efetuado através da determinação do vizinho mais próximo no espaço de faces computado.

4.6.4 Top N Resultados

Após o treino do sistema, a identificação de uma imagem é obtida de forma semelhante para os três algoritmos disponíveis, e pode ser resumida no seguinte conjunto de passos:

1. É criada uma representação de baixo nível da imagem a identificar, aqui designada de prova.
2. A prova é comparada a cada uma das representações de baixo nível criadas durante o treino do sistema e às quais se encontra associado um código numérico representativo da pessoa associada a essa representação.
3. É devolvida o código da pessoa correspondente à menor distância entre a prova e a representação de baixo nível treinada.

Para os algoritmos *Eigenfaces* e *Fisherfaces* a representação de baixo nível de uma imagem corresponde à projeção desta no sub-espço de imagens criado com a análise dos componentes principais, no caso do *Eigenfaces*, e com a posterior análise linear discriminante, no caso do *Fisherfaces*. A comparação entre a prova e as representações de baixo nível criadas no treino é efetuada em ambos os casos através da distância euclidiana entre as matrizes das duas representações.

Para o algoritmo *LBPH* a representação de baixo nível criada corresponde ao histograma representativo das *local binary patterns* de uma imagem. A comparação entre a prova e as representações de baixo nível criadas durante o treino é efetuada através da comparação da distância entre os seus histogramas.

A extensão efetuada à implementação existente na biblioteca *OpenCV*, incide sobre o ponto 3 dos passos acima citados. Em vez de apenas ser devolvido o código identificativo da pessoa cuja representação apresenta menor distância às representações de baixo nível existentes, é devolvida uma lista de n elementos, ordenada de forma crescente pelas distâncias calculadas. Esta lista encontra-se agrupada pelas diferentes entidades, significando assim que cada pessoa apenas aparece uma única vez na lista, independentemente, ou não da existência de várias representações de baixo nível referentes a essa pessoa. Desta forma torna-se possível o desenvolvimento de aplicações que, da uma imagem, respondam às perguntas "Qual é nome das 10 pessoas mais parecidas com a imagem fornecida?" em vez de apenas "Qual é a pessoa mais parecida com a imagem fornecida?".

4.7 Funcionalidades

O trabalho realizado no âmbito desta dissertação resultou na criação de cinco executáveis com diferentes funcionalidades e responsabilidades no âmbito do sistema Visage. Os primeiros três *FaceDetector.exe*, *FaceRecognizer.exe* e *CSVCreator.exe* foram utilizados para efetuar a avaliação reportada no capítulo 5. Já os programas *TrainAndSaveModel.exe* e *TopN.exe* permitem a utilização do sistema Visage como base de outros programas que tirem partido do reconhecimento facial de imagens no seu funcionamento.

FaceDetector.exe Efetua o pré-processamento de uma galeria de imagens de acordo com os seguintes argumentos:

- ***libraryPath.txt***: O ficheiro txt com a descrição das imagens da galeria a processar.
- ***outputDirName***: O caminho do novo diretório a criar.
- ***-mask***: Define se deve ser colocada uma máscara sobre a imagem (ver 4.5.5). Parâmetro opcional.
- ***-N*, *-E* ou *-C***: Define qual o tipo de normalização que deve ser efetuada sobre a imagem, *-N*, *-E* ou *-C* para as técnicas *Contrast Stretching*, Equalização Histograma e CLAHE, respetivamente (ver 4.5.3). Parâmetro opcional.
- ***-G* ou *-B***: Define qual o filtro de abstração a aplicar sobre a imagem, *-G* ou *-B*, para os filtros gaussiano ou bilateral, respetivamente. Para efetuar o pré-processamento com recurso ao filtro kuwahara anisotrópico é necessário recorrer a um programa externo (ver 4.5.4). Parâmetro opcional.
- ***-align***: Define se deve ser efetuado alinhamento das faces detetadas nas imagens (ver 4.5.2). Parâmetro opcional.

FaceRecognizer.exe Efetua a avaliação do desempenho do sistema de reconhecimento facial para uma determinada galeria, de acordo com os seguintes parâmetros:

- ***libraryPath.txt***: O ficheiro txt com a descrição das imagens da galeria a processar.
- ***results.txt***: O ficheiro txt com os resultados da avaliação efetuada num formato legível para humanos.
- ***-E*, *-F* ou *-L***: Define qual o algoritmo de reconhecimento facial a utilizar *-E*, *-F* ou *-L* para *Eigenfaces*, *Fisherfaces* ou *LBPH* (ver 4.6).
- ***nResults*** : Define o número de rankings máximo a avaliar.

CSVCreator.exe Permite criar novas sub-galerias a partir de outras pré-existentes, assim como computar as estatísticas de uma galeria, de acordo com os seguinte parâmetros:

- ***libraryPath.txt***: O ficheiro txt com a descrição das imagens da galeria original.

- **outputfilename**: O nome do ficheiro com a descrição da sub-galeria criada.
- **bottomLimit**: O número de mínimo de imagens por pessoa.
- **topLimit**: O número de máximo de imagens por pessoa.
- **statsFile**: Nome do ficheiro onde ficam armazenadas as estatísticas de uma galeria.

TrainAndSaveModel.exe Permite treinar um modelo de reconhecimento facial de uma galeria passada como argumento e grava-lo para um ficheiro para posterior utilização.

- **libraryPath.txt**: O ficheiro txt com a descrição das imagens da galeria a processar.
- **percentageToTrain**: A percentagem de imagens de cada pessoa a usar como imagem de treino e teste.
- **-E, -F ou -L**: Define qual o algoritmo de reconhecimento facial a utilizar -E, -F ou -L para *Eigenfaces*, *Fisherfaces* ou *LBPH* (ver 4.6).
- **output.xml**: O ficheiro onde deve ser gravado o modelo treinado.

topN.exe Devolve o nomes das n personalidades mais parecidas com a personalidade representada na imagem. Efetua o pré-processamento da imagem passada como argumentos, com aplicação de máscara, normalização através de equalização do histograma e alinhamento da face.

- **imagePath**: O caminho da imagem a reconhecer.
- **model.xml**: O modelo de reconhecimento facial previamente treinado a utilizar.
- **nResults**: O número de elementos máximo a incluir no top de resultados devolvido.

Capítulo 5

Avaliação e Resultados

A hipótese levantada no âmbito desta dissertação é que o uso de abstração em imagens que vão ser alvo de reconhecimento facial pode melhorar o processo de reconhecimento facial automático em imagens. Para a análise desta hipótese foi criado um sistema de reconhecimento facial automático em imagens, assim como, escolhida uma biblioteca de imagens a analisar, tal como descrito no capítulo 4. Neste capítulo, é apresentada a avaliação efetuada ao sistema de reconhecimento facial criado, assim como, os diversos resultados obtidos.

Em primeiro lugar, são apresentados os conjuntos de teste criados para efetuar a avaliação do sistema Visage, a partir da coleção LFW-a, na secção 5.1.

Posteriormente são apresentadas as tarefas de pré-processamento a que as imagens de cada conjunto de teste foram sujeitas, em 5.2, das quais resultou a criação de um conjunto de galerias de imagens pré-processadas.

Em terceiro lugar, são apresentadas as diversas avaliações efetuadas ao sistema, *Closed-set Identification* e *Image Retrieval*, as quais são apresentadas nas secções 5.3 e 5.4, respectivamente. A primeira constitui uma medida padrão de avaliação do desempenho de sistemas de reconhecimento facial automático, a segunda, pretende analisar o desempenho do sistema de um ponto de vista mais próximo de um provável caso de uso, neste caso, *Image Retrieval*. Para além destas duas avaliações, é ainda analisada a variação do desempenho do sistema nos diferentes conjuntos de teste utilizados, a qual se encontra reportada na secção 5.5.

Finalmente, em último lugar, é ainda apresentada uma análise aos resultados obtidos, na secção 5.7.

5.1 Conjuntos de Teste

O desempenho de sistemas de reconhecimento facial pode ser analisado através de diferentes perspetivas, tal como introduzido na secção 2.2 desta dissertação. A biblioteca LFW foi criada com o propósito inicial da análise do sub-problema de reconhecimento facial *pair matching*, no qual o sistema deve decidir se duas faces representam ou não o mesmo indivíduo. No entanto, as características desta biblioteca, nomeadamente no que diz respeito à existência da anotação textual

das pessoas presentes em cada imagem, à existência de apenas uma pessoa representada (de forma relevante) em cada imagem e ainda da normalização do formato das imagens, permitem que esta biblioteca possa ser utilizada com um esforço reduzido para o análise dos restantes sub-problemas de reconhecimento facial automático. As avaliações de desempenho reportadas no âmbito desta dissertação nas secções 5.3, 5.4 e 5.5 não se enquadram no paradigma de *pair matching*, para o qual a biblioteca LFW foi originalmente criada, pelo que se tornou necessário o desenvolvimento de novos conjuntos de teste, os quais passamos a apresentar de seguida.

Na avaliação de sistemas de reconhecimento facial automáticos designam-se de *amostras biométricas* as capturas de características de uma pessoa que permitem efetuar o seu reconhecimento. Dependendo do sistema, as amostras biométricas podem ser apenas uma imagem, um conjunto de imagens, ou um vídeo. No projeto Visage as amostras biométricas de um indivíduo são constituídas por um conjunto de imagens, um exemplo de amostras biométricas de um indivíduo para a biblioteca LFW encontra-se representado nas figuras 5.1 e 5.2. Designam-se ainda de *provas* as amostras biométricas apresentadas ao sistema para serem reconhecidas.

Para a avaliação efetuada foram criados 4 conjuntos de teste. Todos os conjuntos possuem um total de 1180 imagens, correspondentes a amostras biométricas de 59 personalidades distintas, existindo 20 imagens por cada pessoa presente em cada conjunto.

As 20 imagens de cada personalidade incluídas em cada um dos conjuntos de testes, variam conforme o conjunto e foram selecionadas aleatoriamente a partir da ferramenta criada especificamente para a criação de conjuntos de teste, *GaleryCreator.exe*, descrita no capítulo 4.5. Algumas imagens encontram-se presentes em todos os quatro conjuntos de teste criados, enquanto outras encontram-se presentes apenas num único conjunto. No total os quatro conjuntos de teste possuem 1592 imagens diferentes.

Para cada conjunto foram ainda criados dois sub-conjuntos de dados: conjunto \mathcal{G} , designado de galeria de treino e conjunto \mathcal{P} , designado de provas.

A galeria de treino, \mathcal{G} , é constituída por 80% das imagens de cada pessoa (16 imagens por pessoa), sendo as suas imagens utilizadas como amostras biométricas para o treino do sistema de reconhecimento facial. Um exemplo das imagens presentes na galeria de treino para o sujeito Angelina Jolie pode ser visto na figura 5.1.

O conjunto \mathcal{P} contem os restantes 20% das imagens de cada pessoa (4 imagens por pessoa) e, tal como o seu nome indica, as suas imagens são utilizadas para a avaliação do sistema desenvolvido. Um exemplo das imagens presentes nas provas para o sujeito Angelina Jolie pode ser visto na figura 5.2.

As imagens incluídas na galeria de treino e nas provas, diferem também de forma aleatória entre os quatro conjuntos, tendo sido também selecionadas a partir da ferramenta *GaleryCreator.exe*. A criação de 4 conjuntos com igual tamanho mas diferentes imagens de treino e teste tem em vista analisar o desempenho do sistema com diferentes galerias, de forma a determinar se existe uma variação significativa dos resultados obtidos em cada galeria.



Figura 5.1: Exemplo de imagens do conjunto \mathcal{G} para o sujeito Angelina Jolie



Figura 5.2: Exemplo de imagens do conjunto \mathcal{P} para o sujeito Angelina Jolie

5.2 Pré-processamento

O módulo de reconhecimento facial, *Face Recognizer*, disponibilizado pela plataforma *OpenCV*, constituiu uma base sólida para o desenvolvimento do sistema de reconhecimento facial *Visage*, contudo, de modo a tornar este sistema mais completo e versátil, revelou-se necessário adaptá-lo e expandir as suas capacidades, nomeadamente através da aplicação de uma cadeia de pré-processamento às imagens existentes, tal como descrito na secção 4.5 deste relatório. A evolução do sistema desenvolvido foi efetuada de forma gradual e iterativa, e culminou na construção de um conjunto de galerias de imagens, analisadas através de um grupo de experiências que permitem tirar conclusões acerca dos efeitos das diversas etapas de pré-processamento realizadas e da contribuição de cada uma delas para a melhoria do desempenho do sistema criado.

Todas as galerias criadas possuem as imagens contidas nos conjuntos de teste definidos em 5.1, diferenciando-se pelos diferentes passos de pré-processamento a que as imagens foram sujeitas. Na tabela 5.1 encontram-se resumidas as galerias resultantes do conjunto de experiências efetuadas, as quais se encontram descritas pormenorizadamente de seguida.

Tabela 5.1: Galerias criadas após pré-processamento.

Designação	Recortada	Normalização	Filtro Abstração	Máscara	Exemplo
Original	-	-	-	-	3.2a
Cropped	Sim	-	-	-	5.3b
Masked	Sim	-	-	Sim	5.3c
Normalized	Sim	<i>Contrast Stretching</i>	-	Sim	5.3d
Equalized	Sim	Equalização Histograma	-	Sim	5.3e
CLAHE	Sim	<i>CLAHE</i>	-	Sim	5.3f
Bilateral	Sim	Equalização Histograma	Bilateral	Sim	5.3h
Gaussian	Sim	Equalização Histograma	Gaussian	Sim	5.3g
AKF	Sim	Equalização Histograma	Kuwahara Anisotropico	Sim	5.3i

5.2.1 Detecção e segmentação da face

Na revisão efetuada ao estado da arte do reconhecimento facial em imagens destacou-se a importância da resolução de um conjunto de sub-problemas específicos para um reconhecimento facial eficaz (ver 2.2.1), nomeadamente a deteção e segmentação das faces existentes numa imagem e a remoção de elementos de fundo externos à imagem. Para a resolução deste problema, revelou-se necessário a avaliação de diferentes alternativas relativamente à forma como é efetuada a segmentação da face da restante imagem, assim como a determinação do impacto dessa segmentação no processo de reconhecimento.

Após uma análise das diferentes alternativas existentes, e através de alguns testes intermédios realizados durante o período de desenvolvimento, foram então criadas as galerias de imagens

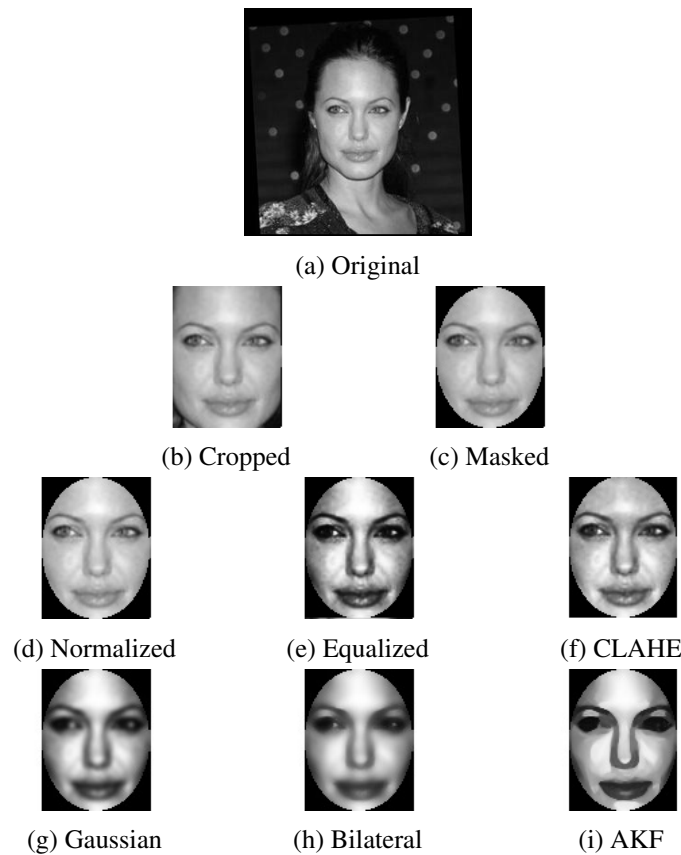


Figura 5.3: Variação da mesma imagem nos vários conjuntos pré-processados analisados.

Original, figura 3.2a; Cropped, figura 5.3b; e Masked, figura 5.3c. A primeira, é constituída pelas imagens originais da biblioteca LFW-a e permite estabelecer uma base de comparação entre as imagens segmentadas e as imagens originais. A galeria cropped é composta por um conjunto de imagens onde as faces foram detetadas e segmentadas pelo detetor facial implementado e descrito no capítulo 4.5. A terceira e última galeria criada possui as mesmas imagens da galeria Cropped, sobre as quais foi posteriormente aplicada uma máscara elíptica de modo a aumentar a área de fundo removida das imagens recortadas.

5.2.2 Normalização do Contraste

A normalização das imagens em termos de contraste é passo comum e essencial à maioria dos sistemas de reconhecimento facial automático modernos, tal como destacado no capítulo 2 desta dissertação. Assim sendo, no decorrer do desenvolvimento do sistema de reconhecimento facial Visage, revelou-se também importante analisar o qual o impacto da normalização do contraste nas imagens utilizadas, assim como qual a melhor forma de efetuar a sua normalização. Foram então estudadas três formas distintas de efetuar a normalização do contraste, *Contrast Stretching*, *Equalização Histograma* e *CLAHE*, descritas em 4.5.3.

Da aplicação da técnica de contrast stretching às imagens presentes nos conjuntos de teste descritos anteriormente resultou a criação da galeria Normalized. As imagens normalizadas através

da equalização simples do histograma encontram-se representadas na galeria Equalized e as normalizadas pela técnica *CLAHE* encontram-se representadas na galeria com o mesmo nome. Para além de normalizadas, as imagens foram também recortadas e após o seu processamento foi aplicada uma máscara elíptica sobre as mesmas. As figuras 5.3d, 5.3e e 5.3f representam exemplos da aplicação das três técnicas avaliadas sobre a mesma imagem.

5.2.3 Abstração Imagens

O último conjunto de galerias criadas visa analisar a hipótese levantada no âmbito desta dissertação, acerca do impacto do uso de filtros de abstração no reconhecimento facial em imagens. A utilização de filtros de abstração constitui uma forma moderna de simplificação do conteúdo visual, permitindo remover informação redundante e aumentar o destaque da mensagem visual a transmitir. Estudos efetuados demonstraram que a aplicação destes filtros na recuperação de informação multimédia, nomeadamente no âmbito da ilustração automática de texto têm a potencialidade de melhorar a informação retornada, assim como reduzir significativamente as necessidades de processamento e armazenamento das imagens [CR12]. O uso de filtros de abstração no âmbito do sistema Visage, tem em vista analisar qual o impacto do uso destes filtros no reconhecimento facial em imagens, quer ao nível do desempenho do sistema, quer ao nível das suas necessidades de armazenamento e processamento das imagens.

Para além de abstraídas as imagens foram também sujeitas às tarefas de pré-processamento que apresentaram melhores resultados nos conjuntos anteriormente criados, nomeadamente a deteção e segmentação da face, a normalização do seu histograma e ainda a aplicação de uma máscara após a aplicação do filtro de abstração.

A abstração das imagens foi efetuada com recurso aos filtros gaussiano, bilateral e kuwahara anisotrópico, descritos em 4.5.4, resultando na criação das galerias Gaussian, Bilateral e AKF, respetivamente.

5.3 Avaliação *Closed-Set Identification*

Tal como introduzido na secção 2.2, o paradigma de *closed-set identification*, constitui um sub-problema de identificação em que uma prova é apresentada ao sistema e pretende-se que este devolva a identidade da pessoa presente na imagem. Neste caso particular do problema de identificação, todas as provas apresentadas possuem uma correspondência na galeria, em oposição ao caso geral de identificação, no qual pode ou não haver uma correspondência.

A avaliação *closed-set* é uma medida padrão de avaliação do desempenho em sistemas de reconhecimento facial automático, tendo sido utilizada em diversas avaliações efetuadas, nomeadamente nas avaliações FERET e FRVT apresentadas na revisão do estado da arte desta dissertação (ver 2). A utilização de um conjunto fechado de imagens permite uma análise detalhada do desempenho de um algoritmo permitindo responder à pergunta "a identificação correta encontra-se nos primeiros n resultados? "em vez de apenas "o primeiro resultado é o correto?".

5.3.1 Metodologia Avaliação

Para cada conjunto de teste descrito em 5.1, seja \mathcal{G} a sua galeria de treino, em que $\mathcal{G} = \{g_1, \dots, g_N\}$ e seja \mathcal{P} o conjunto das suas provas, em que $\mathcal{P} = \{p_1, \dots, p_N\}$. Quando uma prova p_j é apresentada ao sistema, essa prova é então comparada a cada amostra biométrica g_i da galeria, resultando dessa comparação o respetivo índice de similaridade (*similarity score*), s_{ij} . Este índice é designado de *match score*, caso g_i e p_j sejam amostras da mesma pessoa, caso não o sejam é designado de *nonmatch score*. Quanto menor o índice de similaridade, maior é a probabilidade das imagens comparadas pertencerem à mesma pessoa, sendo que o melhor valor possível para o índice de similaridade de um *match score* é zero.

Na identificação de $p_j \in \mathcal{P}$, em primeiro lugar são calculados os índices de similaridade para todas as amostras na galeria \mathcal{G} , sendo posteriormente ordenados os seus resultados. O ranking de p_j , r_{p_j} , é igual a n , se o seu *match score* corresponde ao enésimo menor índice de similaridade. A taxa de identificação para o ranking n , $T_I(n)$, corresponde à fração de provas com ranking n ou menor do que n , ou seja:

$$T_I(n) = \frac{|C(n)|}{|\mathcal{P}|} \times 100 \quad (5.1)$$

Em que $|C(n)|$ é número de provas com ranking n , ou menor do que n , e $|\mathcal{P}|$ o número de provas existentes em \mathcal{P} .

A $T_I(n)$ é calculada para todos os rankings entre 1 e 30. Dentro desse intervalo, os primeiros níveis de ranking apresentam uma maior variação na percentagem de pessoas identificadas, revelando-se assim mais significativos para a análise do desempenho do sistema.

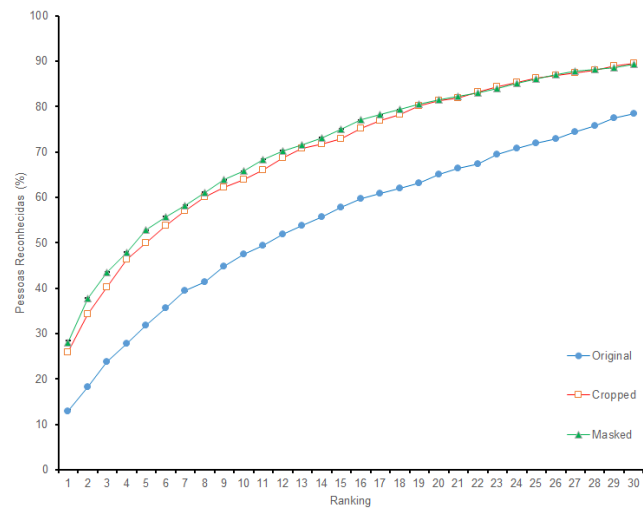
A avaliação é efetuada através da análise do desempenho dos três algoritmos disponíveis em cada uma das galerias de imagens pré-processadas, encontrando-se os resultados obtidos agrupados num conjunto de três experiências, cada uma dedicada a analisar a variação do comportamento do sistema dado um problema específico, utilizando para isso diferentes grupos de galerias de imagens pré-processadas. A divisão entre imagens de treino e teste de cada galeria é efetuada segundo os quatro conjuntos de testes criados, sendo cada galeria avaliada em cada um dos quatro conjuntos de teste existentes e posteriormente calculada a média dos resultados obtidos.

A média dos resultados obtidos encontra-se representada em um gráfico do tipo *Cumulative match score* (CMS). Um gráfico do tipo CMS representa $T_I(n)$ como uma função de ranking n . No eixo horizontal encontra-se representado o ranking e no eixo vertical encontra-se representada a respetiva taxa de identificação, um exemplo de um gráfico CMS pode ser visto em 5.5.

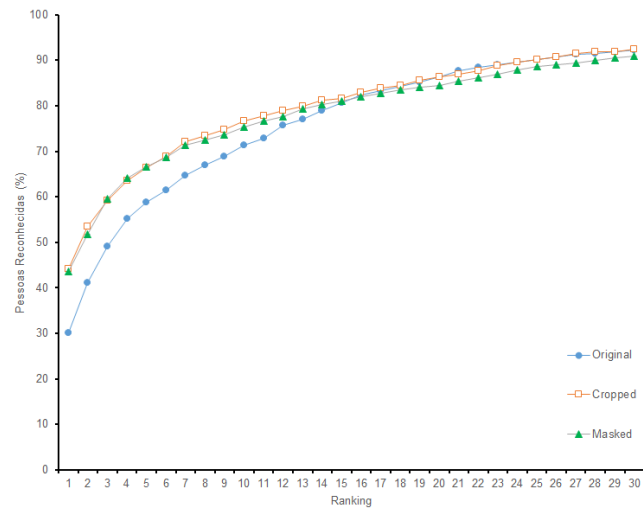
5.3.2 Resultados Experiência 1: Impacto da segmentação das faces

Esta experiência visa analisar o impacto da segmentação das faces presentes na imagem do fundo existente nas mesmas. Os resultados obtidos podem ser visualizados na figura 5.4, onde se encontram ilustrada a performance dos algoritmos, *Eigenfaces*, *Fisherfaces* e *LBPH*, nos três conjuntos criados especialmente para esta experiência, os quais se encontram descritos mais pormenorizadamente em 5.2.1.

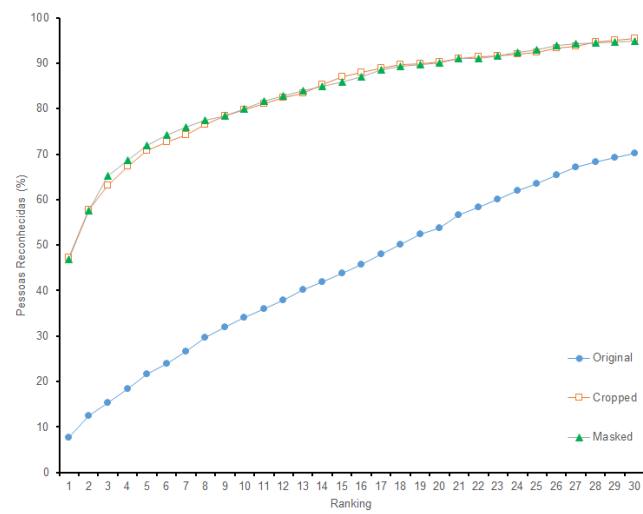
Avaliação e Resultados



(a) Eigenfaces



(b) Fisherfaces



(c) LBPH

Figura 5.4: Desempenho algoritmos Eigenfaces, Fisherfaces e LBPH para os conjuntos Original, Cropped e Masked

Após a análise dos resultados obtidos na experiência 1, nomeadamente através da comparação das taxa de identificação de ranking 1 dos conjuntos Cropped e Masked com o conjunto Original, verifica-se que a segmentação das faces presentes nas imagens originais produz uma melhoria efetiva dos resultados obtidos, traduzindo-se num aumento de cerca de 12%, 14% e 40% para os algoritmos *Eigenfaces*, *Fisherfaces* e *LBPH*, respetivamente.

Uma análise mais aprofundada desses resultados, demonstra ainda que, no caso dos algoritmos *Eigenfaces* e *LBPH*, a diferença na taxa de identificação entre as imagens originais (conjunto Original) e as imagens segmentadas (conjuntos Cropped e Masked) é significativa e relativamente constante nos vários níveis para os quais a taxa de deteção se encontra reportada. Por outro lado, no caso do algoritmo *Fisherfaces*, apesar da diferença significativa no número de imagens corretamente identificadas nos primeiros níveis, a partir do ranking 15, todos os conjuntos de imagens obtêm uma taxa de identificação similar, e com uma progressão semelhante. Esta aproximação no número de caras corretamente identificadas nos rankings mais elevados pode ser explicada pela existência de informação no fundo das imagens, a qual poderá ser utilizada pelo próprio algoritmo para a identificação das mesmas, ao invés da utilização da face contida na imagem, reforçando assim a importância da extração do fundo contido nas imagens originais.

Por outro lado, note-se ainda que, apesar do impacto positivo da segmentação das imagens nos três algoritmos testados, o impacto desta segmentação varia significativamente conforme o algoritmo utilizado, sendo que o *Fisherfaces* é o que revela resultados mais constantes independentemente da manipulação efetuada nas imagens e o algoritmo *LBPH* o que regista uma maior melhoria desses mesmos resultados.

Ao nível da diferença entre as imagens apenas segmentadas (conjunto Cropped) e as imagens com máscara (conjunto Masked), é possível verificar que ambas possuem resultados semelhantes, sendo que a maior diferença foi detetada no algoritmo *Eigenfaces*, onde as imagens com máscara demonstraram um desempenho ligeiramente superior, como é possível verificar pelas $T_1(2)$ de 43,4% e 40,1% para as imagens com máscara e apenas recortadas, respetivamente. Apesar da diferença registada não ser muito significativa, a existência de uma máscara garante uma extração de uma maior quantidade de fundo das imagens, permitindo assim uma maior confiança dos resultados obtidos, no sentido em que se reduz o perigo de identificação pelas características do fundo da imagem e não das características faciais presentes na mesma.

Finalmente, através da análise do gráfico 5.5, é possível concluir que ao nível dos algoritmos utilizados o algoritmo *LBPH* obtém resultados globalmente melhores, ao passo que o algoritmo *Eigenfaces* obtém os piores resultados. Esta diferença verifica-se quer nos rankings inferiores, quer nos superiores, sendo menor quanto maior é o ranking utilizado, como é possível verificar pelas taxas de identificação de ranking 1 de 46,8% e 28,1%, no conjunto Masked, para os algoritmos *LBPH* e *Eigenfaces*, respetivamente, e pelas taxas de identificação de ranking 30 de 94,9% e 89,3%, para os mesmos algoritmos no mesmo conjunto. A análise comparativa do desempenho dos três algoritmos encontra-se reportada no gráfico 5.5 para a galeria masked, uma vez esta foi a galeria que revelou melhores resultados das três analisadas.

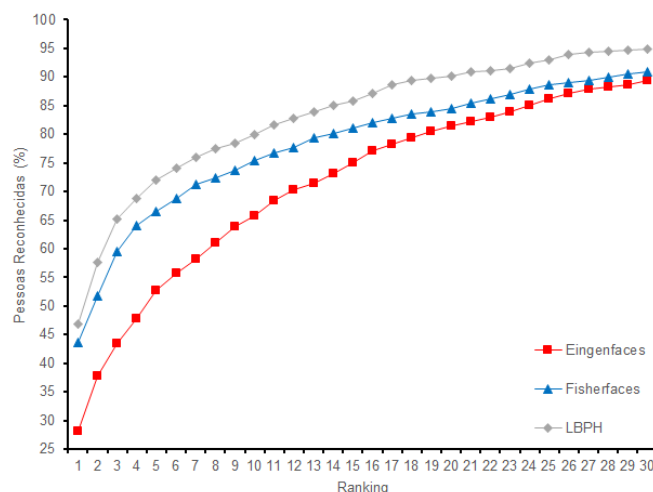


Figura 5.5: Comparação do desempenho dos três algoritmos implementados para o conjunto Masked

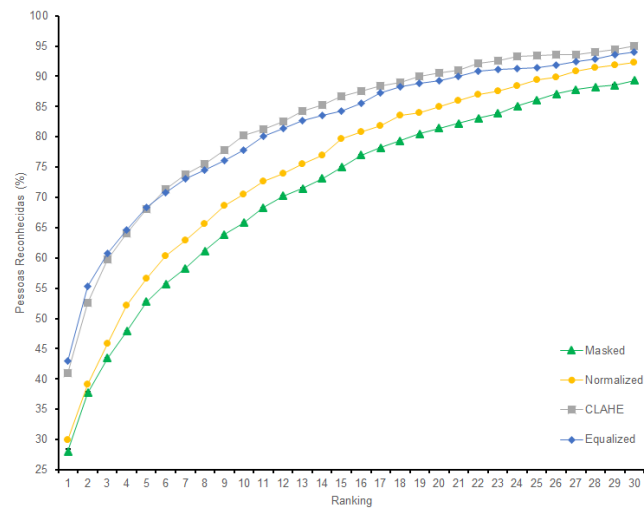
5.3.3 Resultados Experiência 2: Impacto da Normalização do Contraste

A aquisição de imagens em condições não controladas pode resultar na obtenção de fotografias com um contraste reduzido e níveis de iluminação muito distintos. De forma a ultrapassar este problema, a normalização do contraste de uma imagem é uma tarefa comum nos sistemas de reconhecimento faciais modernos. Nesta experiência pretendemos analisar qual o impacto da normalização do contraste de uma imagem nos diferentes algoritmos implementados, assim como determinar qual a melhor estratégia de normalização, de entre três técnicas distintas implementadas, através da análise do desempenho do sistema em três galerias previamente normalizadas com cada uma dessas técnicas (ver 5.2.2).

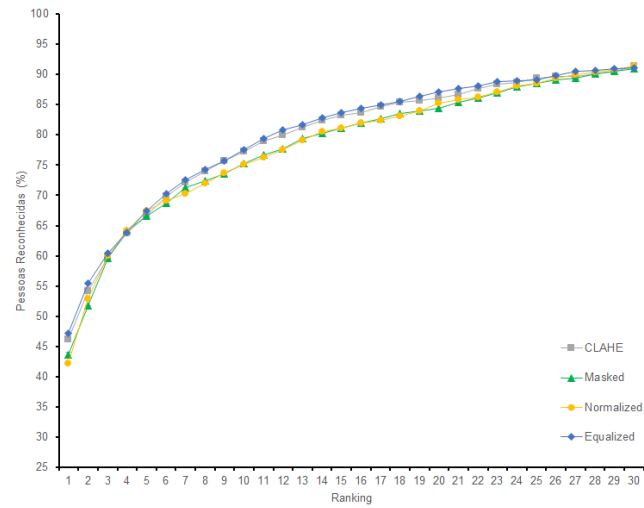
Na figura 5.6 é possível visualizar os resultados para obtidos para o desempenho do sistema nas galerias Normalized, Equalized e CLAHE, correspondentes às técnicas de normalização *contrast stretching*, equalização histograma e CLAHE, respectivamente, as quais encontram-se descritas na secção 4.5.3 deste relatório. Nos gráficos da figura 5.6 encontra-se ainda representados os resultados obtidos para o conjunto Masked, utilizado na experiência 1, o qual serve de referência para a comparação entre os conjuntos não normalizados e os conjuntos normalizados.

Como é possível concluir pela observação da figura 5.6, a normalização das imagens apresenta resultados globalmente melhores do que os obtidos para conjuntos não normalizados. O impacto obtido varia, no entanto, consideravelmente conforme o algoritmo utilizado, sendo que o maior impacto é registado para o algoritmo Eigenfaces, onde existe uma melhoria na ordem dos 15.0% na percentagem de pessoas reconhecidas no primeiro nível do ranking, quando comparadas as galerias Masked e Equalized. Os algoritmos *Fisherfaces* e *LBPH* mostram uma diferença máxima de apenas 3.8% e 2.2% entre os conjuntos Masked e Equalized, pelo que é possível concluir que estes algoritmos possuem uma maior resistência aos efeitos da iluminação no desempenho quando

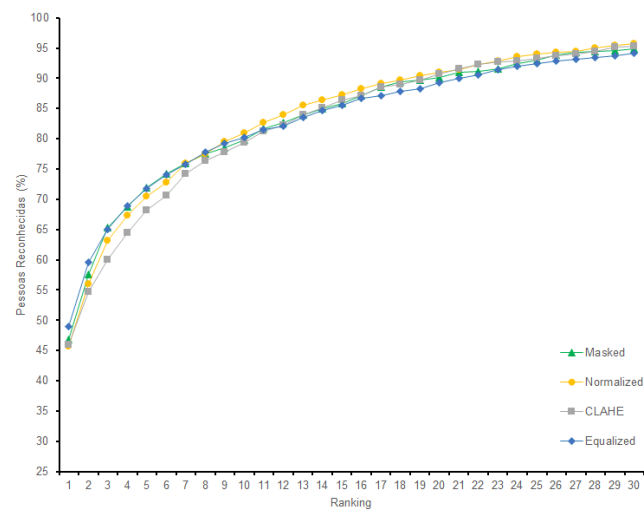
Avaliação e Resultados



(a) Eigenfaces



(b) Fisherfaces



(c) LBPH

Figura 5.6: Desempenho algoritmos Eigenfaces, Fisherfaces e LBPH para os conjuntos Normalized, Equalized e CLAHE

comparados com o algoritmo *Eigenfaces* ¹.

Ao nível das três técnicas utilizadas é possível concluir que as duas variantes de equalização do histograma possuem os resultados mais satisfatórios. A equalização simples do histograma, conjunto Equalized, tem tendência a demonstrar resultados melhores para os primeiros níveis de ranking, sendo posteriormente igualada ou até ultrapassada pela técnica CLAHE nos níveis mais superiores. Uma vez que os primeiros níveis de ranking incluem a informação mais significativa para a maioria dos utilizadores de sistemas de reconhecimento facial automático, consideramos que o desempenho da equalização do histograma das imagens possui resultados mais relevantes para a utilização no sistema de reconhecimento facial Visage do que a técnica *CLAHE*.

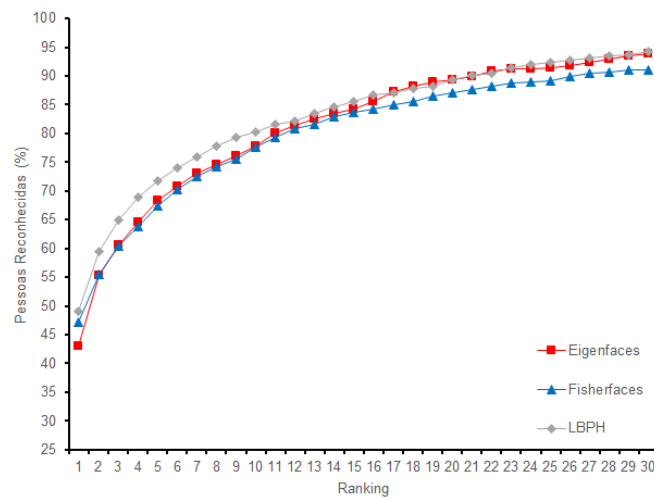


Figura 5.7: Comparação do desempenho dos três algoritmos implementados para o conjunto Equalized

Finalmente, a análise comparativa dos resultados obtidos para os três algoritmos implementados, ilustrada pelo gráfico 5.7, permite concluir que após a normalização do histograma das imagens o desempenho dos algoritmos apresenta uma diferença muito menor do que a registrada anteriormente e ilustrada pelo gráfico 5.5. Nesta experiência o algoritmo LBPH continua a ser o algoritmo com melhor desempenho, registrando no entanto uma diferença de apenas 1.8% e 6.0% para os algoritmos *Fisherfaces* e *Eigenfaces*, respetivamente, na taxa de identificação de ranking 1 no conjunto Equalized. De notar ainda que, a partir do ranking de nível 2, o algoritmo *Eigenfaces* possui uma taxa de identificação idêntica ao valores obtidos pela algoritmo *Fisherfaces*, ultrapassando o desempenho do segundo, a partir do ranking 11. Esta grande evolução no desempenho, notada para o algoritmo *Eigenfaces*, em contraste com a ligeira melhoria obtida pelos restantes denota a complexidade associada ao problema de reconhecimento facial automático, nomeadamente no que diz respeito à forma como diferentes fatores afetam de forma diferenciada diferentes algoritmos e diferentes conjuntos de dados.

¹ A pouca variação obtida neste dois últimos algoritmos vai de acordo com o defendido por investigações anteriores realizadas por [AHP04].

5.3.4 Resultados Experiência 3: Abstração de Imagens

A terceira e última experiência realizada visa analisar qual o impacto da utilização de filtros de abstração no processo de reconhecimento facial em imagens. A análise foi efetuada com recurso a três filtros de abstração diferentes, descritos em 4.5.4, os quais correspondem a diferentes técnicas e níveis de abstração, sendo o filtro gaussiano o filtro com um menor impacto na imagem original e o filtro kuwahara anisotrópico aquele que produz um maior grau de abstração.

Os resultados da análise de desempenho efetuada nesta terceira experiência encontram-se representados na figura 5.8 em quatro galerias distintas. A primeira corresponde à galeria Equalized, utilizada na experiência 2, e que serve como referência de comparação entre as imagens abstraídas e as imagens normalizadas, uma vez que esta foi a galeria onde se registou o melhor desempenho na experiência 2. Da aplicação dos filtros de abstração nas imagens existentes na galeria Equalized, resultou a criação de três galerias distintas Gaussian, na qual foi utilizado o filtro gaussiano para a abstração das imagens, Bilateral, na qual foi utilizado o filtro homónimo para a abstração das imagens e ainda a galeria AKF cujas imagens foram abstraídas com recurso ao filtro kuwahara anisotrópico. Uma descrição mais pormenorizada destas galerias pode ser encontrada em 5.2.3.

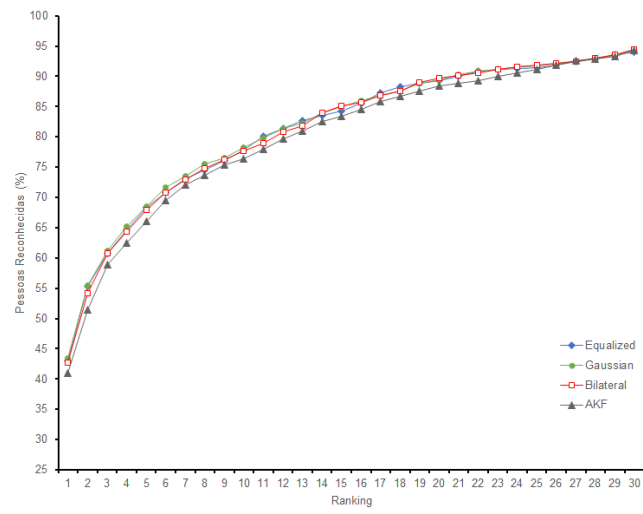
A análise da figura 5.8 permite concluir que, de uma forma global, não se verifica uma melhoria significativa nos resultados obtidos com o recurso à abstração de imagens no desempenho do sistema de reconhecimento facial criado.

O impacto do uso da abstração de imagens varia, no entanto, conforme o algoritmo de reconhecimento utilizado. O algoritmo *Eigenfaces* é o que regista uma menor variação nos resultados obtidos quando comparadas taxas de identificação registadas nas galerias abstraídas com as registadas para a coleção Equalized. A este nível as galerias Gaussian e Bilateral registam um desempenho equivalente ao conjunto de imagens Equalized, enquanto que a galeria AKF regista um desempenho ligeiramente inferior, como é possível verificar pelas taxas de identificação nível 1 de 43.0%, 43.4%, 42.7% para as galerias Equalized, Gaussian e Bilateral e 41.0% para a galeria AKF.

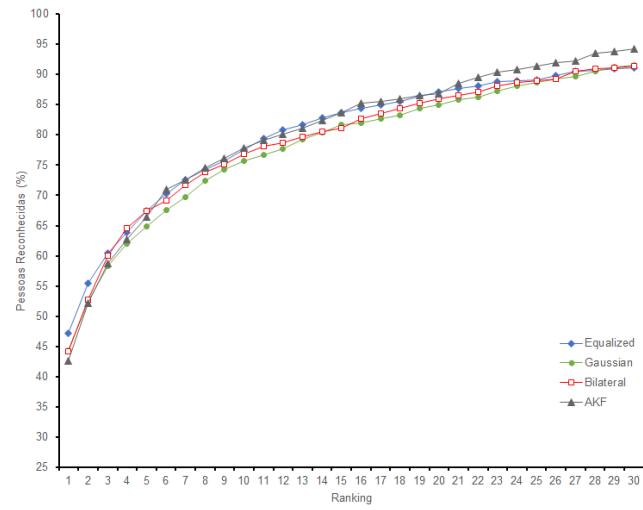
Para o algoritmo *Fisherfaces* a galeria abstraída com recurso ao filtro gaussiano é a que regista um pior desempenho, obtendo uma taxa de identificação inferior, ou equivalente às restantes galerias em todos os níveis para os quais a taxa de identificação se encontra reportada. As galerias Bilateral e AKF, apresentam um desempenho equivalente ao da galeria Equalized, sendo que no caso do conjunto Bilateral o desempenho é regra geral ligeiramente inferior ao das outras duas galerias, ao passo que o o conjunto AKF regista uma melhoria considerável no número de pessoas identificadas a partir do ranking 20 até ao final, como é possível verificar pelas taxas de identificação de ranking 30 de cerca de 91% para os conjuntos Equalized, Gaussian e Bilateral e de 94.2% para o conjunto AKF.

Finalmente, o algoritmo *LBPH* é o que regista um maior impacto na utilização da abstração de imagens no processo de reconhecimento, sendo que nos casos dos conjuntos Bilateral e AKF o desempenho do reconhecimento piora significativamente, registando valores 6.8% inferiores para a taxa de identificação nível 1 no conjunto Bilateral e 17.6% inferiores para o mesmo ranking na

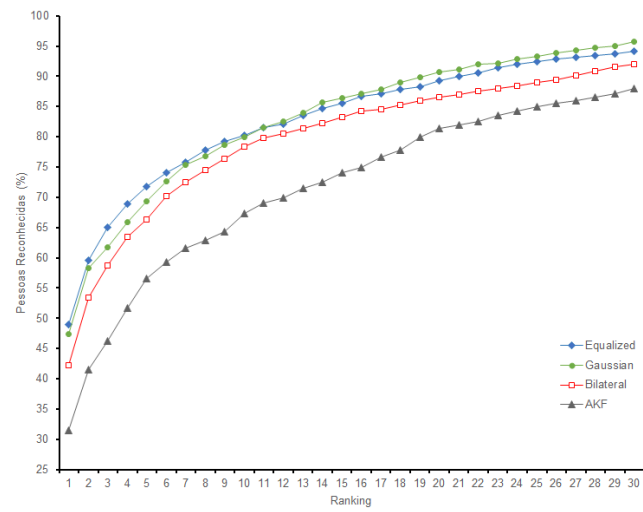
Avaliação e Resultados



(a) Eigenfaces



(b) Fisherfaces



(c) LBPH

Figura 5.8: Desempenho algoritmos Eigenfaces, Fisherfaces e LBPH para os conjuntos Gaussian, Bilateral e AKF

comparação entre o conjunto AKF e Equalized. No caso do conjunto Gaussian, o desempenho é equivalente ou ligeiramente inferior do que o registado para o conjunto Equalized até ao ranking 6, passando a ser ligeiramente superior a partir deste nível até ao final onde é registada uma diferença de 1.5 pontos na percentagem de pessoas identificadas.

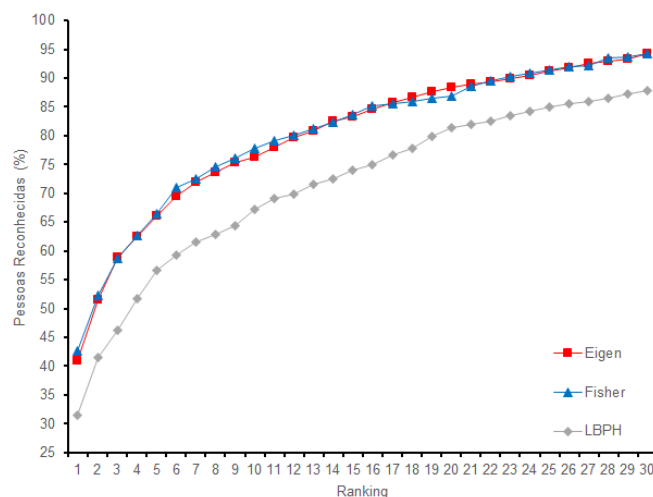


Figura 5.9: Comparação do desempenho dos três algoritmos implementados para o conjunto AKF

5.4 Avaliação *Image Retrieval*

A metodologia de avaliação proposta na secção 5.3 desta dissertação permite-nos efetuar uma avaliação qualitativa do desempenho do sistema de reconhecimento facial criado, assim como dos diferentes algoritmos implementados, através de uma medida padrão para a avaliação de sistemas de reconhecimento facial automáticos. Nesta segunda avaliação propomos a avaliação do desempenho do sistema de um ponto de vista centrado num possível caso de uso do mesmo: *Image Retrieval*.

A área de recuperação de informação multimédia regista atualmente uma importância crescente, resultado do elevado número de conteúdos produzidos, assim como do elevado número de dispositivos de partilha disponíveis, tal como destacado na revisão do estado da arte desta dissertação. Desta forma, a recuperação de informação multimédia e em particular a área de *Image Retrieval* representa uma das áreas com maior potencial para a utilização do sistema de reconhecimento facial criado.

Para uma dada imagem contendo uma face, o sistema Visage é capaz de detetar a face presente na imagem e apresentar uma lista ordenada de possíveis entidades que se encontram representadas nessa imagem, de acordo com um modelo anteriormente treinado. Com base neste sistema é possível desenvolver aplicações de *image retrieval* capazes de encontrar fotos de uma personalidade específica ou apresentar, para uma foto de rosto fornecida pelo utilizador, a celebridade ou figura pública mais parecida.

5.4.1 Metodologia Avaliação

Em recuperação de informação multimédia designa-se de *precisão* a fração de documentos relevantes do total de documentos retornados por uma pesquisa. Ao analisar a precisão podem ser analisados todos os documentos relevantes ou apenas os n documentos mais relevantes, designado-se nessa situação de *precisão a n* .

No sistema de reconhecimento facial criado, para uma dada imagem é apresentada uma lista de possíveis entidades que se encontram representadas nessa imagem, em que o primeiro resultado representa a entidade com maior probabilidade de estar presente na imagem e em que cada entidade aparece uma única vez na lista de resultados possíveis. Caso se pretenda efetuar a identificação de uma personalidade existe apenas um resultado verdadeiramente relevante na lista de resultados obtida, pelo que a medida de *precisão a n* utilizada em recuperação de informação multimédia não é válida para a avaliação efetuada nesta secção.

Para avaliação do desempenho do sistema do ponto de vista de *image retrieval* foi então criada a medida *precisão na galeria* (PG). A *precisão na galeria* resulta da adaptação da precisão de recuperação de informação multimédia ao caso particular do sistema de reconhecimento facial automático desenvolvido. Considerando a definição de ranking de uma prova apresentada em 5.3, a *precisão na galeria* de p_j , PG_{p_j} , corresponde a:

$$PG_{p_j} = \frac{1}{r_{p_j}} \times 100 \quad (5.2)$$

Em que \mathcal{P} é o conjunto de provas da galeria, $p_j \in \mathcal{P}$ e r_{p_j} corresponde ao ranking da prova p_j .

A *precisão na galeria* de uma prova progride então de 100% para ranking 1, 50% para ranking 2, 33% ranking 3 e assim sucessivamente. As provas com um menor ranking têm então uma maior precisão, assim como uma maior distância entre si em termos de precisão (50 unidades entre 1 e 2, 22 unidades entre 2 e 3, e assim sucessivamente). Ora, no caso de *image retrieval* é importante que o resultado mais relevante, a identificação correta de uma pessoa, surja com o menor ranking possível, pelo que a distância nas *precisões na galeria* obtidas nos primeiros resultados deve ser maior do que a obtida para os resultados que surgem posteriormente na lista de entidades possíveis, tal como verificado na métrica adotada. Desta forma, é possível analisar a desempenho do sistema com ênfase nos resultados de topo para uma determinada prova, beneficiando os valores de *precisão na galeria* quando o nome correto aparece cedo na lista de resultado possíveis, ao mesmo tempo que se diferencia significativamente os resultados de topo.

Dada a *precisão na galeria* a *média da precisão na galeria* (MPG) de um algoritmo é calculada através da média da *precisão na galeria* obtida para cada prova, ou seja:

$$MPG_{\text{algoritmo}} = \frac{\sum_{i=0}^n PG_{p_i}}{n} \quad (5.3)$$

Em que n corresponde ao número de provas existentes em \mathcal{P} .

Tabela 5.2: Resultados da Média da Precisão na Galeria (melhores a negrito).

Galeria	$MPG_{Eigenfaces}$	$MPG_{Fisherfaces}$	MPG_{LBPH}
Original	23.1%	43.4%	16.0%
Cropped	37.7%	54.8%	58.2%
Masked	40.0%	54.0%	58.3%
Normalized	42.4%	53.5%	57.2%
Equalized	55.2%	57.3%	59.6%
CLAHE	53.6%	55.9%	56.5%
Gaussian	55.3%	54.2%	58.1%
Bilateral	54.6%	54.7%	53.8%
AKF	52.8%	53.9%	43.0%

Através da *média da precisão na galeria* é possível analisar o desempenho de um algoritmo de uma forma global, ao contrário da avaliação efetuada em 5.3, onde o desempenho de um algoritmo é analisado em função da progressão da sua taxa de identificação em vários rankings.

5.4.2 Resultados

Na tabela 5.2 encontram-se representada a *média da precisão na galeria* dos três algoritmos avaliados em cada uma das galerias introduzidas em 5.2.

Através da análise dos resultados obtidos é possível concluir que o algoritmo *LBPH* é aquele que possui um melhor desempenho a nível global, apresentando os melhores resultados em 6 das galerias avaliadas, ao passo que o algoritmo *Fisherfaces* foi o melhor em apenas 3 das galerias e o algoritmo *Eigenfaces* em nenhuma das galerias.

Por outro lado, o algoritmo *LBPH* revela também uma maior sensibilidade em relação às etapas de pré-processamento efetuadas sobre as imagens, demonstrando a maior variação registada nas nove galerias avaliadas. A este nível, o algoritmo *Fisherfaces* revelou ser o mais robusto ao apresentar uma variação máxima de 13.9% nas várias galerias, enquanto os algoritmos *Eigenfaces* e *LBPH* revelaram uma diferença entre a *média da precisão na galeria* mínima e máxima obtidas nas nove galerias de 32.2% e 43.6%, respetivamente.

Em relação às diversas tarefas de pré-processamento efetuadas o maior impacto é verificado após a deteção da face e respetiva segmentação da restante imagem, sendo que aqui, tal como verificado na experiência reportada em 5.3.3, a aplicação de uma máscara revela-se pouco significativa, apresentando apenas uma ligeira melhoria no desempenho do algoritmo *Eigenfaces*. Comparando os resultados obtidos nas várias galerias é possível concluir que as etapas de pré-processamento aplicadas na galeria Equalized, nomeadamente a segmentação da face, equalização do histograma e aplicação de uma máscara, são as que produzem de forma global um desempenho mais elevado.

Finalmente, o uso de filtros de abstração na cadeia de pré-processamento das imagens produz uma ligeira diminuição da *média da precisão na galeria* obtida, sendo esta mais evidente para

Tabela 5.3: Algoritmos ordenados por desempenho nos diversos conjuntos de teste (L = LBPH, F = Fisherfaces, E = Eigenfaces; melhores a negrito).

Galeria	A	B	C	D
Original	F, E, L	F, E, L	F, E, L	F, E, L
Cropped	L, F, E	L, F, E	L, F, E	L, F, E
Masked	L, F, E	L, F, E	L, F, E	L, F, E
Normalized	L, F, E	F, L, E	L, F, E	L, F, E
Equalized	L, F, E	F, L, E	L, F, E	L, F, E
CLAHE	L, F, E	F, L, E	L, F, E	L, F, E
Gaussian	L, F, E	L, F, E	L, F, E	L, F, E
Bilateral	F, E, L	F, L, E	E, L, F	E, L, F
AKF	E, F, L	F, E, L	F, E, L	E, F, L

algoritmo *LBPH* na galeria *AKF*. As galerias abstraídas registam, no entanto, um desempenho melhor do que as imagens originais e próximo das restantes galerias com imagens pré-processadas.

5.5 Variação Desempenho

Considerando que de um ponto de vista estatístico um algoritmo de reconhecimento facial estima a identidade de uma face, é então possível interrogarmos-nos se, para uma dada categoria de imagens, se verifica uma variação no desempenho de um algoritmo quando são utilizadas galerias de treino e provas diferentes [PRR00].

Os resultados das avaliações efetuadas apresentados nas secções 5.3 e 5.4, representam a média dos resultados obtidos nos quatro conjuntos de teste criados para a avaliação do sistema de reconhecimento facial Visage. Nesta secção é efetuada uma análise à variação dos resultados obtidos para cada um dos conjuntos de teste criados, nas várias galerias e para os três algoritmos de reconhecimento facial utilizados.

Cada um dos conjuntos de teste criados, aqui designados de A, B, C e D, são constituídos por 20 imagens de 59 indivíduos distintos, sendo que a diferença entre os quatro conjuntos se encontra nas imagens utilizadas como galeria de treino e provas, tal com descrito em 5.1. As imagens de cada conjunto foram sujeitas a um conjunto de operações de pré-processamento diferentes, resultantes na criação das galerias Original, Cropped, Masked, Normalized, Equalized, CLAHE, Gaussian, Bilateral e AKF onde cada galeria contém todas as imagens dos quatro conjuntos de teste, tal como descrito em 5.2. A variação do desempenho dos três algoritmos, para cada galeria, nos diversos conjuntos de teste encontra-se reportada nas tabelas 5.3 e 5.4. A primeira permite analisar se o desempenho de um conjunto em relação aos restantes varia conforme as etapas de pré-processamento aplicadas sobre as imagens, assim como, verificar se existe variação na performance de um algoritmo em relação aos restantes conforme a galeria utilizada. A segunda tabela,

Tabela 5.4: Desvio padrão (σ) entre os valores de MPG obtidos para os 4 conjuntos de teste (melhores a negrito).

Galeria	$\sigma_{Eigenfaces}$	$\sigma_{Fisherfaces}$	σ_{LBPH}
Original	1.21%	1.12%	0.39%
Cropped	2.44%	2.41%	1.91%
Masked	2.83%	3.29%	1.46%
Normalized	2.09%	3.85%	1.36%
Equalized	1.13%	2.26%	1.13%
CLAHE	2.38%	4.03%	1.17%
Gaussian	1.80%	2.22%	2.51%
Bilateral	1.54%	3.03%	1.92%
AKF	1.26%	3.93%	2.43%

5.4, reporta o desvio padrão (σ) existente entre os resultados obtidos para os quatro conjuntos de teste nas nove galerias.

Tal como é possível verificar através da observação da tabela 5.3, no conjunto B foram obtidos os melhores resultados em todas as galerias analisadas, verificando-se assim que o desempenho relativo do sistema num conjunto em relação aos restantes não é significativamente afetado pelas etapas de pré-processamento aplicadas nas imagens.

Por outro lado, na tabela 5.3, verifica-se a existência de variação ao nível do algoritmo com melhor desempenho conforme o conjunto de teste analisado, como demonstra o facto de no conjunto B o algoritmo *Fisherfaces* obter os melhores resultados em seis das nove galerias analisadas, enquanto que nos restantes conjuntos o algoritmo *LBPH* regista o melhor desempenho na maioria das galerias. A existência desta variação é ainda reforçada pela existência de galerias onde o algoritmo com melhor desempenho varia conforme o conjunto analisado, como é visível na galeria Equalized.

Finalmente, a tabela 5.3 permite ainda confirmar que o algoritmo com melhor desempenho num dado conjunto de teste, varia conforme as etapas de pré-processamento aplicadas sobre as imagens desse conjunto, como é possível verificar ao analisar o algoritmo como melhor desempenho nas galerias Gaussian, Bilateral e AKF no conjunto de teste A.

Por último, ao nível do desvio padrão registado entre os quatro conjuntos e sintetizado na tabela 5.4, verifica-se que este varia desde 0.39% para o algoritmo *LBPH* no conjunto Original, até um máximo de 4.03% para o algoritmo *Fisherfaces* no conjunto CLAHE. A este nível, destaca-se o facto de o algoritmo com melhor desempenho na maioria dos conjuntos de teste, *LBPH*, ser também o que apresenta um menor desvio padrão na maioria das galerias, fator ainda mais

Tabela 5.5: Necessidades de armazenamento galerias pré-processadas.

Galeria	Total(MB)	Médio (KB)	Máximo (KB)	Mínimo (KB)	Diferença
Original	125.5	81	114	47	-
Cropped	20.9	13	18	9	-83.4%
Masked	19.4	12	16	9	-84.5%
Normalized	20.7	13	16	10	-83.5%
Equalized	23.7	15	18	12	-81.1%
CLAHE	22.9	15	18	12	-81.7%
Gaussian	22.4	14	17	12	-82.2%
Bilateral	21.7	14	16	11	-82.7%
AKF	13.7	9	11	6	-89.1%

evidenciado se as galerias abstraídas não forem consideradas.

5.6 Análise Necessidade Armazenamento

O desempenho de um sistema de reconhecimento facial automático pode ser analisado do ponto de vista da sua performance no reconhecimento, tal como efetuado em 5.3 e 5.4, no entanto, as necessidades de processamento e armazenamento das imagens utilizadas por este tipo de sistemas devem também ser objeto de análise para a avaliação do mesmo.

Na tabela 5.5, encontram-se sintetizadas as necessidades de armazenamento das imagens que constituem as galerias utilizadas na avaliação do sistema Visage. Todas as galerias pré-processadas reduzem em pelo menos 81% as necessidades de armazenamento necessárias, sendo que a este nível se destacam as imagens da galeria AKF, abstraídas com recurso ao filtro kuwahara anisotrópico, e que são em média 89% inferiores às imagens originais. As imagens abstraídas traduzem-se assim em imagens mais fáceis e rápidas de processar pelo sistema, ao mesmo tempo que permitem a redução da quantidade de dados transferidos na utilização do sistema Visage em cooperação com eventuais clientes web ou aplicação móveis.

5.7 Discussão Resultados Obtidos

Capítulo 6

Conclusões e Perspetivas Futuras

6.1 Conclusões

O projeto de dissertação apresentado neste relatório técnico tem como objetivo estudar o impacto do uso de abstração de imagens no âmbito do reconhecimento facial automático. Para isso, pretendemos criar um sistema de reconhecimento facial de personalidades que integre a abstração de imagens no processo de reconhecimento facial. Este sistema deverá ser capaz de detetar as faces presentes numa imagem e apresentar uma lista de possíveis entidades nela contidas.

Através da análise do estado arte apresentada é possível concluir que o reconhecimento facial em imagens é um tema atual e onde se tem verificado um interesse crescente devido às suas múltiplas áreas de aplicação, assim como ao elevado valor comercial tradicionalmente associado a este tipo de soluções.

O problema de reconhecimento facial é, no entanto, um problema complexo que integra um conjunto de sub-problemas complexos. Estes sub-problemas, aliados as múltiplas áreas de aplicação do reconhecimento facial, fazem com que exista uma grande variação do desempenho dos diversos sistemas existentes, a qual se encontra diretamente relacionada com as condições de utilização dos mesmos, nomeadamente ao no que diz respeito às galerias de imagens utilizadas. A este nível, em situações onde as condições de captura das imagens são controladas e existe uma cooperação ativa por parte dos utilizadores os resultados obtidos são muito satisfatórios, sendo mesmo considerado que o problema se encontra praticamente resolvido. Em contraste, em situações de captura não controladas e onde exista uma variação da iluminação, pose e expressão dos indivíduos este é ainda um problema desafiante e onde se verifica necessidade de investigação e melhoria dos resultados obtidos.

Os filtros de abstração de imagens constituem uma forma moderna e computacionalmente eficaz de abstração de informação, sendo tradicionalmente utilizados para comunicar mais eficazmente uma mensagem visual. O uso destes filtros para a pesquisa baseada em conteúdos com vista a ilustração automática de texto demonstrou resultados positivos, pelo que a abordagem apresentada neste relatório de utilização dos mesmos filtros no processo de reconhecimento facial tem a potencialidade de apresentar resultados igualmente positivos.

Os resultados obtidos no âmbito desta dissertação permitirão contribuir para o desenvolvimento da área através da utilização de uma nova abordagem ao processo de reconhecimento facial automático, assim como tirar conclusões relevantes acerca do uso destes filtros no processo de reconhecimento. Adicionalmente a implementação do sistema com base em uma biblioteca de código aberto permite também alargar o número de soluções atualmente existentes a este nível. Finalmente, o uso da biblioteca Sapo Fama, permitirá ainda uma valorização do projeto do ponto de vista comercial, uma vez que esta pode ser uma das suas áreas diretas de aplicação.

6.2 Perspectivas Futuras

Referências

- [AHP04] T Ahonen, A Hadid, and M Pietikäinen. Face recognition with local binary patterns. *Computer Vision-ECCV 2004*, pages 469–481, 2004.
- [AHP06] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. Face description with local binary patterns: application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12):2037–41, December 2006.
- [BAC09] Soma Biswas, Gaurav Aggarwal, and Rama Chellappa. Robust estimation of albedo for illumination-invariant matching and shape recovery. *IEEE transactions on pattern analysis and machine intelligence*, 31(5):884–99, May 2009.
- [BBP01] Duane M. Blackburn, Mike Bone, and P. J. Phillips. Face Recognition Vendor Test 2000: Evaluation Report. page 71, February 2001.
- [BHK97] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.
- [BK08] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. O’Reilly Media, Inc., 2008.
- [BT54] Jerome S. Bruner and Renanto Tagiuri. THE PERCEPTION OF PEOPLE. page 61, September 1954.
- [BV03] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, September 2003.
- [Chi] Chinese Academy Of Sciences. Facial recognition technology safeguards Beijing Olympics. http://english.cas.cn/Ne/CASE/200808/t20080815_18764.shtml. Acedida: 11/01/13.
- [CR12] Filipe Coelho and Cristina Ribeiro. Image abstraction in crossmedia retrieval for text illustration. In *Proceedings of the 34th European conference on Advances in Information Retrieval, ECIR’12*, pages 329–339, Berlin, Heidelberg, 2012. Springer-Verlag.
- [CSP10] Rama Chellappa, Pawan Sinha, and P. Jonathon Phillips. Face Recognition by Computers and Humans. *Computer*, 43(2):46–55, February 2010.
- [DC93] G Deng and L W Cahill. An adaptive Gaussian filter for noise reduction and edge detection. In *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record.*, pages 1615–1619 vol.3, 1993.

REFERÊNCIAS

- [EC97] Kamran Etemad and Rama Chellappa. Discriminant analysis for recognition of human face images. *Journal of the Optical Society of America A*, 14(8):1724, August 1997.
- [Fis36] R. A. Fisher. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(2):179–188, September 1936.
- [GQP10] Patrick J Grother, George W Quinn, and P Jonathon Phillips. Report on the Evaluation of 2D Still-Image Face Recognition Algorithms. *National Institute of Standards and Technology*, 7709, 2010.
- [HRBLm07] Gary B Huang, Manu Ramesh, Tamara Berg, and Erik Learned-miller. Labeled Faces in the Wild : A Database for Studying Face Recognition in Unconstrained Environments. Technical report, University of Massachusetts, Amherst, 2007.
- [Int] Intel Corporation. What Happens In An Internet Minute? <http://www.intel.com/content/www/us/en/communications/internet-minute-infographic.html>. Acedida: 25/11/12.
- [its] itseez - Opencv Developers Team. ABOUT | OpenCV. <http://opencv.org/about.html>. Acedida: 19/01/13.
- [Kan73] Takeo Kanade. *Computer recognition of human faces*. Birkhauser, Basel, 1973.
- [KCWI12] Jan Eric Kyprianidis, John Collomosse, Tinghuai Wang, and Tobias Isenberg. State of the ‘Art’: A Taxonomy of Artistic Stylization Techniques for Images and Video. *IEEE Transactions on Visualization and Computer Graphics*, 2012.
- [KKD09] Jan Eric Kyprianidis, Henry Kang, and Jürgen Döllner. Image and Video Abstraction by Anisotropic Kuwahara Filtering. *Computer Graphics Forum*, 28(7):1955–1963, October 2009.
- [KL08] Henry Kang and Seungyong Lee. Shape-simplifying Image Abstraction. *Computer Graphics Forum*, 27(7):1773–1780, October 2008.
- [KLC09] Henry Kang, Seungyong Lee, and Charles K Chui. Flow-based image abstraction. *IEEE transactions on visualization and computer graphics*, 15(1):62–76, January 2009.
- [KS90] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, 1990.
- [LJ11] SZ Li and AK Jain. *Handbook of face recognition*. Springer, 2011.
- [LKP03] Rainer Lienhart, Alexander Kuranov, and Vadim Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. *Pattern Recognition*, 2003.
- [Mina] Ministério da Administração Interna. Passaporte Electrónico Português. <http://www.pep.pt/estatisticas.html>. Acedida: 10/01/13.
- [Minb] Ministério da Administração Interna. RAPID - Reconhecimento Automático de Passageiros Identificados Documentalmente. = <http://www.rapid.sef.pt/>. Acedida: 10/01/13.

REFERÊNCIAS

- [paMaA] Agência para a Modernização Administrativa. Estatísticas Cartão Cidadão. http://www.cartaodecidadao.pt/index.php?option=com_content&task=view&id=295&Itemid=114&lang=pt. Acedida: 05/01/13.
- [PPC07] Giuseppe Papari, Nicolai Petkov, and Patrizio Campisi. Artistic Edge and Corner Enhancing Smoothing. *IEEE Transactions on Image Processing*, 16(10):2449–2462, October 2007.
- [PRR00] P.J. Phillips, S.A. Rizvi, and P.J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.
- [PSO07] PJ Phillips, WT Scruggs, and AJ O’Toole. FRVT 2006 and ICE 2006 large-scale results, 2007. *National Institute of Standards and Technology*, (March), 2007.
- [PWHR98] P.Jonathon Phillips, Harry Wechsler, Jeffery Huang, and Patrick J. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295–306, April 1998.
- [Rez04] Ali M. Reza. Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement. *The Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology*, 38(1):35–44, August 2004.
- [Ser82] J Serra. *Image analysis and mathematical morphology*. London.: Academic Press.[Review by Fensen, EB in: J. Microsc. 131 (1983) 258.] Cell size, Staining Microscopy Technique, Mathematics, General article Review article (PMBD, 185707888), 1982.
- [TM98] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pages 839–846. Narosa Publishing House, 1998.
- [TP91] Matthew Turk and Alex Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, January 1991.
- [VB] Vision-Box. Vision-Box. <http://www.vision-box.com/>. Acedida: 11/01/13.
- [VJ01] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 1:I–511–I–518, 2001.
- [VJ04] Paul Viola and Michael J. Jones. Robust Real-Time Face Detection. *International Journal of Computer Vision*, 57(2):137–154, May 2004.
- [WFKvdM97] L Wiskott, J M Fellous, N Kuiger, and C von der Malsburg. Face recognition by elastic bunch graph matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):775–779, 1997.
- [WOG06] Holger Winnemöller, Sven C. Olsen, and Bruce Gooch. Real-time video abstraction. *ACM Transactions on Graphics*, 25(3):1221, July 2006.

REFERÊNCIAS

- [YKMA02] Ming-hsuan Yang, David J Kriegman, Senior Member, and Narendra Ahuja. Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.
- [ZCK98] W. Zhao, R. Chellappa, and A. Krishnaswamy. Discriminant analysis of principal components for face recognition. In *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 336–341. IEEE Comput. Soc, 1998.
- [ZCPR03] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A Literature Survey. *ACM Computing Surveys*, 35(4):399–458, December 2003.