

Reconhecimento Facial Através de Aprendizagem Profunda em Imagens Multiespectrais

Luis Lopes Chambino
Academia Militar, Lisboa
Instituto Superior Técnico (IST),
Universidade de Lisboa
chambino.lcl@exercito.pt

José Silvestre Silva
Academia Militar, Lisboa
CINAMIL, Lisboa
LIBPhys-UC, Coimbra
jose.silva@academiamilitar.pt

Alexandre Bernardino
Instituto Superior Técnico (IST),
Universidade de Lisboa
Instituto de Sistemas e Robótica (ISR),
Portugal
alex@isr.tecnico.ulisboa.pt

Resumo – O reconhecimento facial é um método de identificação ou autenticação da identidade de indivíduos através dos seus rostos. Atualmente, os sistemas de reconhecimento facial que utilizem imagens multiespectrais obtêm melhores resultados, em comparação com aqueles que utilizem apenas imagens da banda espectral do visível. Neste trabalho, é proposto um detetor de pele a ser aplicado num módulo de deteção de falsificação e uma arquitetura que utiliza múltiplas redes neurais convolucionais profundas e imagens multiespectrais para efetuar o reconhecimento facial. É realizado um estudo com o objetivo de avaliar o desempenho da adaptação de diversas camadas da rede neural base. Adicionalmente, foi realizado um segundo estudo para avaliar o desempenho das Máquinas Vetoriais de Suporte (SVM) e dos classificadores k-Nearest Neighbor para classificar os conjuntos de características multidimensionais obtidas através da arquitetura proposta. Os resultados experimentais nas bases de dados multiespectrais Tufts e CASIA NIR-VIS 2.0 indicam um desempenho competitivo no reconhecimento facial obtendo uma pontuação de *rank-1* de 99,7% e 99,8%, respetivamente.

I. INTRODUÇÃO

Nos dias de hoje, muitos dos sistemas biométricos de reconhecimento facial funcionam na banda espectral do visível. Estes possuem diversos obstáculos, tais como: oclusões, variação de poses, cooperação da pessoa e, o mais problemático, alterações na luminosidade. Como resultado, é necessário complementar os atuais sistemas de reconhecimento facial, com a utilização de outros sensores biométricos (e.g., impressão digital ou íris) ou com outras bandas espectrais, a fim de minimizar estes problemas [1] [2].

Na Tabela 1 estão indicadas as bandas espectrais mais utilizadas e aplicadas em reconhecimento facial multiespectral.

TABELA 1. INTERVALOS ESPECTRAIS UTILIZADOS EM RECONHECIMENTO FACIAL [4].

Nome da Banda Espectral	Comprimento de Onda (μm)
Visível	0,38 – 0,75
<i>Near Infrared</i> (NIR)	0,75 – 1,40
<i>Short Wavelength Infrared</i> (SWIR)	1,40 – 3,00
<i>Mid Wavelength Infrared</i> (MWIR)	3,00 – 8,00
<i>Long Wavelength Infrared</i> (LWIR)	8,00 – 15,00

A utilização do espectro eletromagnético infravermelho, nomeadamente as bandas espectrais *Near Infrared* (NIR), *Short Wavelength Infrared* (SWIR), *Mid Wavelength Infrared* (MWIR) e *Long Wavelength Infrared* (LWIR), têm sido utilizadas com sucesso em sistemas de reconhecimento facial, como complemento do espectro visível [1] [3]. A estes sistemas, que utilizam mais do que uma banda espectral, denominam-se de multiespectrais.

A utilização do espectro eletromagnético infravermelho em sistemas de reconhecimento facial possui diversas vantagens quando comparada com o espectro eletromagnético do visível. O infravermelho é impercetível ao olho humano e, ao mesmo tempo, menos sensível às diferenças de luminosidade. Por exemplo, as câmaras noturnas utilizadas na vigilância por vídeo utilizam *LEDs*, com emissão no espectro infravermelho de forma a iluminar o local, e realizar vigilância noturna sem que as pessoas tenham conhecimento [5].

Dado que as bandas espectrais NIR e SWIR estão próximas da banda espectral do visível, é possível adaptar os métodos de aprendizagem automática treinados com imagens do espectro visível. As bandas espectrais MWIR e LWIR (também conhecida por térmica) permitem a utilização de sistemas de reconhecimento facial à noite, quando a luminosidade é reduzida ou mesmo nula.

Este trabalho está organizado da seguinte forma: na Secção II é efetuado o estudo do estado da arte sobre métodos de reconhecimento facial multiespectral, das métricas mais utilizadas e das bases de dados multiespectrais públicas; na Secção III é definida e proposta a metodologia para a deteção facial multiespectral; na Secção IV são indicadas as bases de dados multiespectrais utilizadas, apresentados os resultados obtidos e a sua respetiva análise e discussão; a Secção V é efetuada a conclusão tendo por base os resultados e discussão apresentados na secção anterior, e ainda sugeridos possíveis trabalhos futuros.

II. ESTADO DA ARTE

Nesta secção é apresentada uma revisão sistemática do reconhecimento facial multiespectral, sendo descritos de forma breve os métodos utilizados.

Para efetuar a revisão sistemática foi utilizado a base de dados *Web of Sciences*. Foram considerados artigos publicados de 2000 até junho de 2020 em revistas

internacionais com fator de impacto; trabalhos publicados em conferências não foram considerados. O produto da pesquisa foram 283 artigos publicados em 132 revistas internacionais científicas com fator de impacto.

Para este estado da arte foi necessário efetuar uma segunda seleção mais refinada. Como tal, foram considerados apenas os artigos que efetuavam reconhecimento facial ou deteção facial com duas ou mais bandas espectrais (e.g., VIS-NIR, VIS-LWIR, VIS-NIR-LWIR, NIR-LWIR, entre outras possíveis combinações), reduzindo assim o número de artigos para 47. Estes artigos foram considerados os mais relevantes para a análise em questão.

Foi realizada uma análise destes artigos tendo em consideração as bases de dados multiespectrais e métricas de avaliação utilizadas, tendo sido concluído que a base de dados mais utilizada é a CASIA NIR-VIS 2.0 [6], a qual foi empregue 15 vezes nos artigos analisados.

A métrica mais utilizada para comparar resultados entre metodologias, quando utilizada a mesma base de dados, foi o *rank-1*.

Através da análise sistemática foi possível destacar cinco métodos principais utilizados no reconhecimento facial multiespectral: reconhecimento por características, subespaço acoplado, síntese, fusão e redes neurais profundas. A metodologia mais utilizada corresponde às redes neurais profundas, tendo sido empregue por 32% dos trabalhos relevantes analisados.

Na Figura 1 está ilustrado a distribuição dos métodos de reconhecimento facial por ano de publicação. Da análise deste gráfico, é possível concluir a predominância dos artigos que utilizam o método de fusão de imagens em reconhecimento facial multiespectral até 2016 inclusive. A partir de 2017, o método de fusão foi superado pelo método de redes neurais profundas, uma vez que este produz resultados superiores.

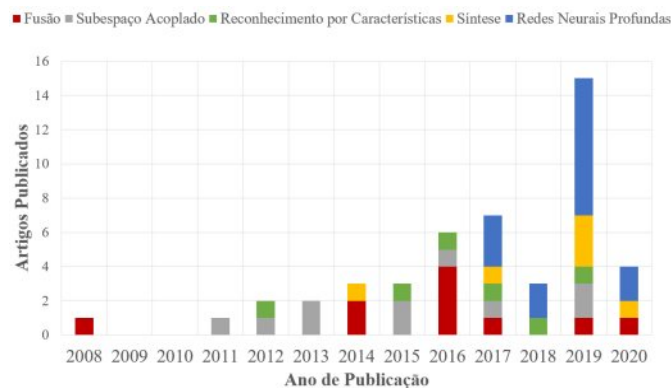


Fig. 1. Distribuição dos métodos por ano de publicação.

A Figura 2 apresenta um gráfico caixa de bigodes que resume o desempenho obtido nos artigos estudados para cada método. Tal estudo permite comparar o nível de desempenho em cada método.

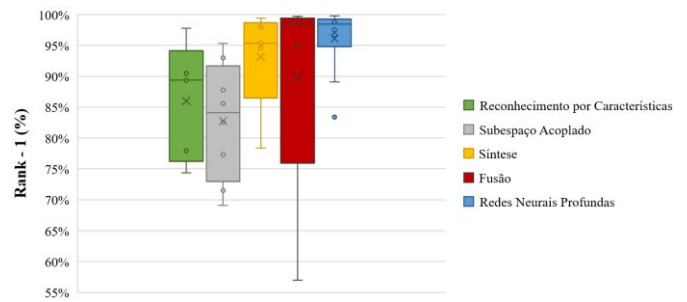


Fig. 2. Desempenho obtido nos métodos descritos nos artigos analisados.

Como é possível observar pela Figura 2 as redes neurais profundas obtêm os melhores resultados, justificando a procura e criação de novas redes neurais profundas, sendo esta área de investigação muito ativa (também corroborado pela Figura 1, na coluna ano 2019).

III. METODOLOGIA

Neste capítulo é apresentada a metodologia adotada para a implementação de um sistema de reconhecimento facial multiespectral. Na Figura 3 está ilustrado o fluxograma da metodologia proposta.

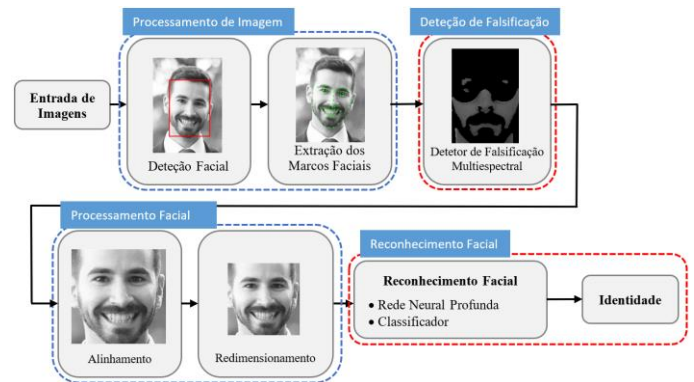


Fig. 3. Fluxograma da metodologia proposta.

A metodologia proposta inicia-se pela obtenção de imagens multiespectrais (i.e., visível e infravermelho).

O passo seguinte é a conversão das imagens para a gama de cores do cinzento, deteção das faces humanas e a extração dos marcos faciais (e.g. olhos, nariz, boca) na imagem. O módulo encarregue por efetuar esta tarefa é o módulo de processamento de imagem.

O sistema de reconhecimento facial proposto inclui um módulo para detetar e avisar potenciais ataques de falsificação, denominado de módulo de deteção de falsificação. Este módulo tira partido de todas as imagens multiespectrais disponíveis para efetuar a deteção da pele, prevenindo assim o sistema de reconhecimento facial de possíveis ataques.

No módulo seguinte é realizado o processamento facial, onde são utilizados os marcos faciais, obtidos no primeiro módulo (i.e., processamento de imagem), de forma a alinhar a face. Este módulo tem como principal objetivo a

normalização da imagem a inserir na rede neural convolucional profunda (DCNN).

O módulo de reconhecimento facial, extrai um conjunto de características da pessoa a identificar. De seguida, é feita uma classificação da identidade mediante o conjunto de características obtidos anteriormente de forma a obter a identificação da pessoa nas imagens multiespectrais processadas inicialmente no primeiro módulo.

A principal vantagem da arquitetura proposta é permitir a utilização de diversos canais, alocando para cada canal uma banda espectral, ou intervalo espectral (caso se esteja a utilizar vários intervalos espectrais na mesma banda).

A DCNN base utilizada em cada canal é a LightCNN [7]. A rede base possui como entrada imagens na gama de cores do cinzento, com tamanho de 128x128 pixéis, e como saída, um conjunto de características, representativo da identidade da pessoa, de 256 dimensões. São adaptadas camadas distintas da LightCNN [7] de forma a adequar o modelo utilizado a uma banda espectral diferente do visível.

Tendo como ponto de partida a LightCNN, foi adicionada no fim da rede uma nova camada, denominada de última camada ligada (UCL), tendo como entrada um conjunto de características com dimensões de $N \times 256$, que através de uma transformação linear produz um conjunto de características de 256-dimensões.

Na Figura 4 está ilustrado um caso genérico de utilização da rede proposta que utiliza N canais, estando assinalado a verde as camadas que são adaptadas, e a azul as camadas que não são adaptadas. O *Canal 0*, atribuído à banda espectral do visível, não é adaptado. Ao reutilizar-se os pesos de uma DCNN já pré treinada para reconhecimento facial numa base de dados com um número elevado de imagens faciais, é evitado um possível sobre ajuste, dado o número limitado de imagens multiespectrais utilizadas na fase de treino [8].

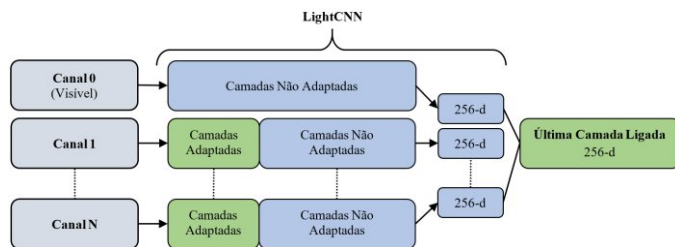


Fig. 4. Esquema da arquitetura da DCNN proposta.

IV. RESULTADOS E DISCUSSÃO

Nesta secção serão descritas as bases de dados multiespectrais utilizadas, os testes efetuados com a finalidade de aferir qual as camadas a serem adaptadas e quais os classificadores que devem ser utilizados.

A. Bases de Dados Multiespectrais

De forma a avaliar corretamente os algoritmos incluídos na metodologia proposta foram utilizados três bases de dados multiespectrais: a Tufts [9], a CASIA NIR-VIS 2.0 [6] e um conjunto de imagens adquiridas na Academia Militar (AM).

A base de dados Tufts [9] é composta por três bandas espectrais, VIS NIR e LWIR. Após a sua limpeza esta ficou com um total de 7 675 imagens faciais de 109 pessoas, tendo sido excluído 53 imagens faciais e 4 pessoas.

A CASIA NIR-VIS 2.0 [6] é composta por duas bandas espectrais, VIS e NIR, constituída por 17 489 imagens faciais de 715 pessoas.

A base de dados elaborada na AM é composta por três bandas espectrais, VIS, SWIR e LWIR. Esta base de dados foi contruída para ser utilizada durante a fase de experimentação do módulo de deteção de falsificação. Durante a construção da base de dados foram utilizadas diversas máscaras, com o intuito de simbolizar diferentes ataques de falsificação. Na Figura 5 estão ilustradas algumas das imagens presentes na base de dados multiespectral elaborada na AM.



Fig. 5. Exemplo de imagens da base de dados multiespectral construída na Academia Militar.

B. Deteção de Falsificação

No módulo de deteção de falsificação foi empregue um classificador de pele que tira proveito de todas as bandas espectrais disponíveis de forma a efetuar uma correta classificação de pele.

O classificador efetua uma classificação ao nível do pixel. Numa primeira etapa é determinada a diferença normalizada, $d[g_a, g_b]$, para todas as combinações possíveis de imagens faciais, tendo em conta os canais disponíveis, usando a expressão:

$$d[g_a, g_b] = \left(\frac{g_a - g_b}{g_a + g_b} \right) \quad (1)$$

em que g corresponde ao valor da intensidade do pixel para o canal a e b , com $1 \leq a \leq n$ e $a \leq b \leq n$, onde n corresponde ao número de canais disponíveis no módulo de deteção de falsificação. A diferença normalizada resulta em valores de $-1 \leq d[g_a, g_b] \leq +1$.

Obtida a diferença normalizada é possível agora aplicar o classificador de pele que utiliza os valores da diferença normalizada de forma a classificar os pixéis como pele ou não pele.

A gama de valores escolhida para classificar como pele ou não pele, foi definida empiricamente. Os valores para pele estão compreendidos entre:

$$(76, 51, 65) < (d[g_r, g_g], d[g_r, g_b], d[g_g, g_b]) < (131, 140, 127)$$

Na Figura 6 está ilustrado o “antes” (imagem à esquerda) e o “produto final” após a aplicação do classificador proposto (imagem à direita) durante a utilização de uma máscara na zona ocular. Nesta figura, a região a preto corresponde ao que foi classificado como não sendo pele.



Fig. 6. O resultado final após aplicação do classificador proposto. Imagem original (esquerda) e imagem após classificador (direita).

Para validar a nossa proposta, foram utilizados outros dois classificadores, o YCbCr e HSV. Na avaliação foi tido em consideração o número de marcos faciais, obtidos no módulo de processamento de imagem, que foram incorretamente classificados como pele. Após análise dos resultados experimentais, é possível afirmar que os classificadores de pele que utilizam a banda espectral do visível tem dificuldades em diferenciar a pele real (i.e., pele humana) de pele falsificada, proveniente das máscaras de falsificação utilizadas.

Quando utilizados estes classificadores de pele, do visível, como método de deteção de falsificação, a taxa de deteção de falsificação (i.e., quando este consegue detetar corretamente um ataque de falsificação) foi de 13%. Resultado justificado pelo facto de os classificadores utilizados não serem capazes de efetuar uma correta diferenciação da pele humana de a máscara.

Quando comparados com o classificador que utiliza as três bandas espectrais, este obtém melhores resultados. A taxa de deteção de falsificação para este foi de 83%. Comparativamente com os classificadores anteriores, este consegue efetuar uma correta diferenciação da máscara, como é possível observar na Figura 6.

C. Reconhecimento Facial

Depois de efetuado o processamento das imagens faciais das bases de dados (i.e., deteção e alinhamento facial), é necessário extrair o conjunto de características dessas mesmas imagens, a fim de classificar a identidade da pessoa presente nas imagens.

Foram efetuados diversos testes para averiguar quais as camadas mais adequadas para o treino da rede neuronal; que valores devem ser tomados nos hiperparâmetros dos classificadores; e, após escolha dos melhores hiperparâmetros, determinar o melhor classificador para classificar o conjunto de características.

As imagens de cada base de dados foram divididas em três conjuntos: treino, validação e teste. A percentagem de

imagens para o conjunto de treino foi de 64%, para o conjunto de validação 16%, e por fim, 20% para o conjunto de teste. Esta partição de cada base de dados foi uma divisão estratificada de forma a que cada pessoa tem um número equitativo de imagens faciais suas em cada subconjunto.

1. Procedimento de Treino

Foi utilizado o alargamento de dados para obter um modelo mais generalizado. Nas imagens presentes no conjunto de dados de treino foi utilizado espelhagem aleatória horizontal e um recorte aleatório para redimensionar a imagem numa resolução de 128x128 pixéis (as imagens no início do treino da rede possuem uma resolução de 144x144 pixéis). Para o conjunto de validação foi apenas aplicado um recorte ao centro, para cumprir o requerimento de resolução da LightCNN [7].

2. Camadas Adaptadas

Foi efetuado um teste para estimar quais as camadas adequadas a adaptar na arquitetura da LightCNN [7]. Inicialmente é adaptado apenas a UCL. Esta camada inicialmente não existia na inicial da DCNN, tendo sido implementada posteriormente para que a saída da arquitetura continuasse a ser o conjunto de características de 256-dimensões.

De seguida, as camadas iniciais da DCNN são adaptadas (inclusive a UCL) até que todas as camadas da DCNN sejam adaptadas. Em todas as experiências os pesos foram inicializados a partir do modelo inicial da LightCNN [7].

A nomenclatura dos grupos adaptados segue a nomenclatura inicial dos autores da LightCNN [7]. As notações utilizadas na combinação das camadas adaptadas são as seguintes: (i) UCL – apenas a UCL é adaptada, (ii) Conv1-N-UCL ({1-N}+UCL) – adapta as camadas de 1 a N em conjunto com a UCL, e (iii) Todas as Camadas – todas as camadas da LightCNN e a UCL são adaptadas.

O número de épocas durante o procedimento de treino foi de 10 e 30 para as bases de dados Tufts [9] e CASIA NIR-VIS 2.0 [6], respetivamente. Após o treino foram extraídos os conjuntos de características de 256-dimensões das imagens de cada base de dados multiespectral.

Efetuada o treino da DCNN com a adaptação de diversas camadas é possível afirmar que, independentemente da base de dados utilizada, para a métrica *rank-1* obtiveram-se os valores de 99,7% e de 99,8% para as bases de dados Tufts [9] e CASIA NIR-VIS 2.0 [6], respetivamente, para o conjunto das camadas ({1-3}+UCL).

3. Estudo dos Hiperparâmetros

Com o melhor modelo, obtido na secção anterior, foi efetuada a extração dos conjuntos de características de 256-dimensões. Para classificar estes conjuntos foram utilizados os classificadores SVM (com *kernel* linear e RBF) e kNN.

Foi aplicada a validação cruzada estratificada, efetuada 5 vezes para a base de dados Tufts [9] e 4 vezes para a base de dados CASIA NIR-VIS 2.0 [6].

O hiperparâmetro afinado para o classificador SVM-Linear foi o parâmetro de regularização (C) de $10^{-10} \leq C \leq 10^{+5}$ com incrementos de uma década.

Para o classificador SVM-RBF foram afinados os seguintes hiperparâmetros: o parâmetro de regularização (C), de $10^{-4} \leq C \leq 10^{+7}$, e o coeficiente de *kernel* (γ) de $10^{-10} \leq \gamma \leq 10^{+2}$, tendo sido utilizados incrementos de uma década para ambos.

Por último, para o classificador kNN, o hiperparâmetro a afinar foi o número de vizinhos próximos (k). A gama de valores analisada para o hiperparâmetro k foi de $1 \leq k \leq 25$, com incrementos de uma unidade.

Após o estudo observa-se que os classificadores SVM, independentemente do *kernel* utilizado, obtêm a pontuação em *rank-1* superior, obtendo uma pontuação em *rank-1* de 99,89% e 99,86% para as bases de dados Tufts [9] e CASIA NIR-VIS 2.0 [6], respetivamente. É também observável que, o melhor valor para o número de vizinhos, hiperparâmetro da kNN, é igual, independentemente da base de dados multiespectral utilizada.

4. Comparação com o Estado da Arte

Efetuada o estudo dos hiperparâmetros mais adequados para cada classificador, é necessário avaliar o desempenho de cada classificador para o conjunto de teste de cada base de dados multiespectral. Simultaneamente, é efetuado um estudo para diferentes classificações através da utilização de uma curva características de correspondência cumulativa (CMC) para auxiliar na decisão do classificador. Nesta curva são traçados a taxa de identificação no eixo das ordenadas e a *rank-N* no eixo das abcissas.

A partir da utilização dos valores em *rank-1* e dos valores obtidos para diferentes classificações (i.e., curva CMC) é possível aferir qual o melhor classificador para cada base de dados multiespectral.

A metodologia proposta utiliza a DCNN LightCNN [7] como base, adaptando as camadas ($\{1-3\} + \text{UCL}$), para produzir um conjunto de características de 256-dimensões, que são classificadas posteriormente pelo classificador SVM-Linear.

Resultados usando a Base de Dados Tufts

Após avaliação do conjunto de teste, da base de dados Tufts, com os classificadores (i) SVM-Linear, (ii) SVM-RBF e (iii) kNN, foram obtidos os seguintes valores em *rank-1*: (i) 99,7%, (ii) 99,2% e (iii) 99,2%.

Na Figura 7 está ilustrado a curva CMC, para os três classificadores, para as primeiras dez classificações (i.e., rank-10) para a base de dados Tufts [9].

Realizando uma análise comparativa entre os três classificadores através dos valores em *rank-1*, o classificador SVM-Linear é aquele que obtém os melhores resultados, com uma pontuação em *rank-1* de 99,7%. Comparativamente, os classificadores SVM-RBF e kNN obtiveram ambos o resultado de 99,2%, para o mesmo conjunto de imagens.

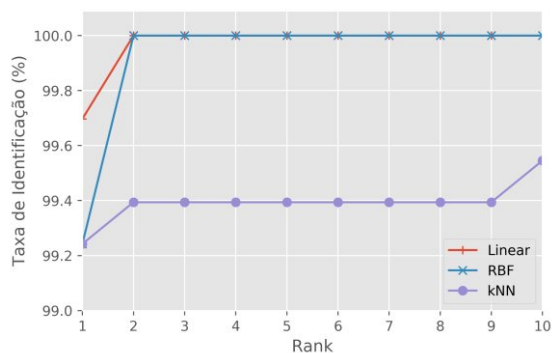


Fig. 7. Curva CMC para os classificadores SVM-Linear, SVM-RBF e kNN, para o conjunto de teste da base de dados multiespectral Tufts.

Na Figura 7 é possível constatar que, para o *rank-2* e independentemente do *kernel* utilizado no classificador SVM, este obtém uma taxa de identificação de 100%. Ou seja, a totalidade das imagens faciais do conjunto de teste foram corretamente identificadas. Comparativamente, o classificador kNN apenas consegue obter uma taxa de identificação de 100% no *rank-10*.

Na Tabela 2 apresenta-se os resultados obtidos pela metodologia proposta e por outras metodologias descritas na literatura. A negrito está salientado o método que obteve melhor pontuação em *rank-1*.

TABELA 2. RESULTADOS OBTIDOS ATRAVÉS DA METODOLOGIA PROPOSTA QUANDO COMPARADOS COM O ESTADO DA ARTE PARA A BASE DE DADOS TUFTS.

Método	Rank-1
TR-GAN [10]	88,7 %
Circular HOG [11]	94,5 %
Metodologia Proposta	99,7 %

A Tabela 2 mostra que a metodologia proposta produz um resultado bastante competitivo em comparação com os resultados obtidos por outras metodologias. De realçar que, sendo a base de dados recente, disponibilizada ao público para investigação em 2020, o número de autores que a utilizam ainda é reduzido.

Resultados usando a Base de Dados CASIA NIR-VIS 2.0

Depois de processado o conjunto de teste da base de dados CASIA NIR-VIS 2.0, com os classificadores (i) SVM-Linear, (ii) SVM-RBF e (iii) kNN, foram obtidos os seguintes valores em *rank-1* de (i) 99,8%, (ii) 99,8% e (iii) 99,7%.

Na Figura 8 está ilustrado a curva CMC, para os três classificadores, as primeiras dez classificações para a base de dados CASIA NIR-VIS 2.0.

Numa análise comparativa entre os três classificadores através dos valores de *rank-1*, ambos os classificadores SVM, independentemente do *kernel* utilizado, obtém uma pontuação de 99,8%. Em comparação, o classificador kNN obtém uma pontuação em *rank-1* de 99,7%.

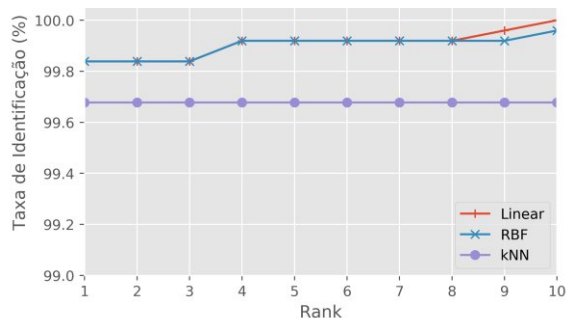


Fig. 8. Curva CMC para os classificadores SVM-Linear, SVM-RBF e kNN, para o conjunto de teste da base de dados CASIA NIR-VIS 2.0.

A Figura 8 mostra que a partir *rank-8* o classificador SVM-Linear obtém uma taxa de identificação superior, quando comparado com o SVM-RBF. Os classificadores SVM-Linear e SVM-RBF obtém uma taxa de identificação de 100% para os *rank-10* e *rank-12*, respetivamente.

Na Tabela 3 encontram-se indicados os resultados obtidos da metodologia proposta e de outras metodologias descritas no estado arte.

TABELA 3. RESULTADOS OBTIDOS ATRAVÉS DA METODOLOGIA PROPOSTA QUANDO COMPARADOS COM O ESTADO DA ARTE PARA A BASE DE DADOS CASIA NIR-VIS 2.0.

Método	Rank-1
LightCNN [7]	96,7 %
DSU [8]	96,3 %
CycleGAN [12]	99,4 %
Metodologia Proposta	99,8 %

Observando a Tabela 3 verifica-se que a metodologia proposta obtém resultados superiores em *rank-1* quando comparado com os valores de outros métodos descritos no estado da arte. O trabalho mais recente que utiliza a base de dados CASIA NIR-VIS 2.0 é o artigo de Bae *et al.* [12] que obteve uma pontuação em *rank-1* de 99,4%, inferior ao resultado obtido pela metodologia proposta, de 99,8%.

Após uma análise detalhada da Tabela 3, é possível observar que a metodologia base da LightCNN [7] obteve uma pontuação em *rank-1* de 96,7%. Através da metodologia proposta foi possível melhorar significativamente a pontuação em *rank-1* de +3,1%.

V. CONCLUSÕES

Neste trabalho, é proposta uma nova arquitetura para o reconhecimento facial utilizando imagens multiespectrais. A arquitetura produz conjuntos de características de 256-dimensões que representam a identidade de uma pessoa baseada em imagens multiespectral.

Adicionalmente, neste estudo é proposto um detetor de personificação com o objetivo de detetar a utilização de máscaras. O classificador multiespectral obtém melhores resultados quando comparado com o YCbCr ou o HSV.

Foram realizados diversos testes com o intuito de identificar quais as melhores camadas a adaptar da LightCNN

e concluiu-se que o melhor resultado corresponde ao conjunto das camadas ($\{1-3\}+UCL$). Este estudo permitiu concluir que quanto maior for o número de camadas adaptadas, pior é o resultado final. Os melhores resultados ocorrem da adaptação das camadas iniciais da rede neural.

Para classificar os conjuntos de características de 256-dimensões o classificador SVM com *kernel* linear obteve os melhores valores em *rank-1* quando comparado com os restantes classificadores, para as duas bases de dados multiespectrais utilizadas.

Estudos extensivos nas bases de dados multiespectrais demonstraram a superioridade da metodologia proposta, tendo sido obtidos valores em *rank-1* de 99,7% e 99,8% para as bases de dados multiespectrais Tufts e CASIA NIR-VIS 2.0.

AGRADECIMENTOS

Este trabalho foi apoiado, em parte, pelo Centro de Investigação, Desenvolvimento e Inovação da Academia Militar (CINAMIL) no âmbito do projeto Reconhecimento Facial Multi-Espectral, e pela FCT com o projeto LARSyS - FCT Project UIDB/50009/2020.

REFERÊNCIAS

- [1] Jain, A. Ross, and K. Nandakumar, Introduction to Biometrics. Springer, 2011.
- [2] R. Munir and R. Khan, "An Extensive Review on Spectral Imaging in Biometric Systems: Challenges and Advancements," Journal of Visual Communication and Image Representation, vol. 65, no. 1, p. 14–26, 2019.
- [3] W. Zhang, X. Zhao, J. Morvan, and L. Chen, "Improving Shadow Suppression for Illumination Robust Face Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 3, pp. 611–624, 2019.
- [4] A. D'Amico, C. Natale, F. Castro, S. Iarossi, A. Catini, and E. Martinelli, "Volatile Compounds Detection by IR Acousto-Optic Detectors," in Unexploded Ordnance Detection and Mitigation. Springer Netherlands, 2009, pp. 21–59.
- [5] S. Hu, N. Short, B. Riggan, M. Chasse, and M. Sarfraz, "Heterogeneous Face Recognition: Recent Advances in Infrared-to-Visible Matching," in International Conference on Automatic Face Gesture Recognition. Washington DC, USA: IEEE, 2017, pp. 883–890.
- [6] S. Li, D. Yi, Z. Lei, and S. Liao, "The CASIA NIR-VIS 2.0 Face Database," in IEEE Conference on Computer Vision and Pattern Recognition Workshops. Portland, United States of America: IEEE, 2013, p. 348–353.
- [7] X. Wu, R. He, Z. Sun, and T. Tan, "A Light CNN for Deep Face Representation with Noisy Labels," IEEE Transactions on Information Forensics and Security, vol. 13, no. 11, pp. 2884–2896, 2018.
- [8] T. D. Pereira, A. Anjos, and S. Marcel, "Heterogeneous Face Recognition Using Domain Specific Units," IEEE Transactions on Information Forensics and Security, vol. 14, no. 7, pp. 1803–1816, 2019.
- [9] K. Panetta, et al., "A Comprehensive Database for Benchmarking Imaging Systems," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 3, pp.509–520, 2020.
- [10] L. Kezbeou, V. Oludare, K. Panetta, and S. Aгаian, "TR-GAN: Thermal to RGB Face Synthesis with Generative Adversarial Network for Cross-Modal Face Recognition," in Mobile Multimedia/Image Processing, vol. 11399. SPIE, 2020.
- [11] S. Rajeev, K. Shreyas, Q. Wan, K. Panetta, and S. Aгаian, "Illumination Invariant NIR Face Recognition Using Directional Visibility," Electronic Imaging, Image Processing: Algorithms and Systems, pp. 273–1–273–7, 2019.
- [12] H. Bae, T. Jeon, Y. Lee, S. Jang, and S. Lee, "Non-Visual to Visual Translation for Cross-Domain Face Recognition," IEEE Access, vol. 8, no. 7, pp. 50 452–50 464, 2020.