

# Age-Invariant Face Recognition to Aid Visually Impaired People

Gonçalo Ferreira Neto<sup>1</sup>, Rodrigo Veras<sup>1</sup>, Kelson Aires<sup>1</sup>, Laurindo Britto Neto<sup>1</sup> \*

<sup>1</sup>Departamento de Computação – Universidade Federal do Piauí (UFPI)  
Teresina – PI – Brazil

g.neto13@hotmail.com, {rveras, kelson, laurindoneto}@ufpi.edu.br

**Abstract.** *This paper proposes an approach to assist people with visual impairments in recognizing people regardless of age. The goal is to develop a system that uses a face recognition approach, focusing on age invariance, that returns good results compared to the results obtained in the literature review. The approach studied uses deep Convolutional Neural Networks - CCNs, pre-trained by the VGGFace2 dataset, to extract feature descriptors from face images and classify with the Linear SVM classification algorithm. As can be seen throughout the paper, the approach returned 89.9% accuracy using the FG-NET dataset with 1002 images. And using the CACD dataset, which contains 163,446 images divided into four different subsets, three sets for training and one for testing, the approach returned 85.2%, 82.4%, and 88.2% accuracy for each model trained with a different training set.*

**Resumo.** *Este trabalho propõem uma abordagem para auxiliar pessoas com deficiência visual no reconhecimento de pessoas independente da idade. O objetivo é desenvolver um sistema que utilize uma abordagem de reconhecimento facial, com foco na invariância na idade, que retorne bons resultados comparados aos resultados obtidos na revisão da literatura. A abordagem estudada utiliza Redes Neurais Convolucionais profundas - CCNs, pré-treinadas pelo conjunto de dados VGGFace2, para extrair descritores de características de imagens de faces e classificar com o algoritmo de classificação Linear SVM. Como pode ser visto no decorrer do trabalho, a abordagem retornou 89,9% de acurácia, utilizando o conjunto de dados FG-NET, com 1002 imagens. E utilizando o conjunto de dados CACD, que contém 163.446 imagens divididas em quatro subconjuntos diferentes, três conjuntos para treino e um para teste, a abordagem retornou 85,2%, 82,4% e 88,2% de acurácia para cada modelo treinado com um conjunto de treinamento diferente.*

## 1. Introdução

Pessoas com deficiência visual possuem muitas dificuldades em seu dia a dia e uma delas é o reconhecimento de pessoas ao seu redor. Para reconhecer um familiar, os indivíduos cegos, geralmente, recorrem à voz ou ao seu cheiro [Britto Neto 2016]. Os indivíduos cegos, que participaram de um dos experimentos realizado por Britto Neto [Britto Neto 2016], contaram várias histórias de suas vidas diárias, associadas à problemas por causa da falta

---

\*Este trabalho tem o apoio da Fundação de Amparo a Pesquisa do Piauí, por meio do Edital FAPEPI/MCT/CNPq N° 007/2018 (PPP) convênio FAPEPI/CNPq.

de consciência em relação à presença ou ausência de pessoas ao redor. Por exemplo, eles relataram que quando estão conversando com alguém, muitas vezes, a pessoa com quem estão conversando sai do ambiente em que se encontram sem avisá-los e os deixa falando sozinhos ou, até mesmo, ela se distraí, usando algum aplicativo de celular (e.g., de redes sociais ou de mensagens), e para de respondê-los, fazendo com que eles fiquem com a dúvida sobre a presença dessa pessoa.

Para solucionar esse problema foi desenvolvido em [Britto Neto et al. 2015] uma abordagem de reconhecimento facial *wearable* que alcançou altas taxas de acurácia e, ao mesmo tempo, economizou recursos de hardware como bateria, memória e processamento. Além disso, ela se mostrou eficiente em situações adversas de aquisição de imagens, como pose, iluminação, *background*, expressões faciais e alguns disfarces. Entretanto, o estudo realizado não procurou contornar o problema do envelhecimento. Resolver esse problema seria de grande auxílio para pessoas com deficiência visual, pois elas não precisariam atualizar o sistema de reconhecimento facial presente no dispositivo *wearable*, periodicamente, com o cadastro das imagens de pessoas do seu convívio.

O Reconhecimento Facial Invariante na Idade (*Age-Invariant Face Recognition – AIFR*) ainda é um desafio para os estudos atuais na área de visão computacional. Tem como um de seus objetivos obter altas taxas de acurácia no reconhecimento de uma mesma pessoa em diferentes idades. No entanto, não é só a questão da idade que dificulta os estudos, existem outros fatores que influenciam na sua robustez, como as condições de aquisição das imagens, pose, iluminação, expressões, *background*, obstrução (i.e., uso de disfarces, óculos escuros, chapéu etc.), entre outros.

As mudanças em um indivíduo podem ser relacionadas ao aparecimento de rugas, mudanças na textura da pele, ganho de peso, perda de peso, crescimento de pelos no rosto, entre outros fatores. Tudo isso acaba dificultando que os sistemas de reconhecimento faciais identifiquem um indivíduo ao longo do tempo.

O objetivo desta pesquisa é chegar a um método que realize AIFR, alcançado altas taxas de acurácia e que futuramente possa ser implementado em um sistema *wearable*, para que as pessoas com deficiência visual possam usar esse dispositivo vestível e tenham a capacidade de realizar o reconhecimento de pessoas próximas sem a necessidade de falar diretamente com o indivíduo. Para isso, foram realizados experimentos para avaliação do desempenho da abordagem proposta sobre o tema, com os conjuntos de dados mais usados e disponíveis encontrados na revisão da literatura. Tal método será incorporado ao sistema *wearable* desenvolvido em [Britto Neto et al. 2015, Britto Neto et al. 2017], por meio de uma arquitetura cliente-servidor, o cliente sendo um dispositivo *wearable* e o servidor estando na nuvem, de forma que o processamento mais custoso computacionalmente seja realizada no servidor, economizando recursos de hardware do dispositivo cliente.

Este artigo está dividido em mais quatro seções: na Seção 2 é apresentado os resultados obtidos de uma revisão da literatura realizada; na Seção 3 são descritos os materiais utilizados, os conjuntos de dados, a abordagem proposta e sua metodologia experimental; na Seção 4 são apresentados e discutidos os resultados obtidos; e por fim, na Seção 5 é concluído o trabalho e são previstas expectativas de trabalhos futuros.

## 2. Revisão da Literatura

Existem várias pesquisas sobre reconhecimento facial, mas a maioria delas não leva em conta a invariância na idade. De acordo com Yang et al. [Yang et al. 2014], os trabalhos realizados nessa área seguem duas categorias: Abordagens Generativas (AGs) e Abordagens Discriminativas (ADs). As AGs tentam gerar imagens de como a pessoa seria na idade requisitada, usando imagens 2D ou modelos 3D para criar o efeito de envelhecimento e, ao final, comparar a imagem gerada e a imagem de teste. Porém, as AGs podem ser muito custosas computacionalmente, e podem gerar suposições de faces que estejam longe da realidade. Já as ADs retiram características discriminativas que não são afetadas ao longo do tempo e, em seguida, usam essas características para classificar os indivíduos. Tais abordagens, também, não são tão confiáveis, pois elas precisam de imagens em várias faixas etárias para conseguir computar características invariantes a idade.

Existem outras abordagens que utilizam Redes Neurais Convolucionais (*Convolutional Neural Network* – CNN), que neste trabalho foram denominadas de Abordagens CNNs (ACNNs). Para aprender as características invariantes na idade, as CNNs podem ser treinadas com outros conjuntos de dados, que contenham imagens de faces em diferentes faixas etárias. Em seguida, podem ser testadas em um conjunto de dados específico. Esse tipo de abordagem requer um conjunto de dados de grande escala. Por isso, a maioria dos trabalhos encontrados na literatura, que desenvolveram ACNNs, realizaram seus testes nas bases MORPH Album 2 [Ricanek and Tesafaye 2006] e CACD [Chen et al. 2015]. Tais conjuntos de dados possuem um grande número de imagens.

Zhao et al. [Zhao et al. 2018], ao contrário das técnicas atuais que extraem características variáveis de idade para reconhecimento ou primeiro sintetizaram uma face que corresponde à idade alvo antes da extração de características, eles propuseram que é mais desejável executar ambas as técnicas para se obter um melhor resultado, constituindo uma abordagem, que neste trabalho serão chamadas de Abordagens Híbridas (AHs).

### 2.1. Abordagens Discriminativas (ADs)

Sungatullina et al. [Sungatullina et al. 2013] desenvolveram o *Multiview Discriminative Learning* (MDL). Inicialmente, a abordagem extrai três diferentes descritores de características locais: *Scale Invariant Feature Transform* (SIFT), *Local Binary Patterns* (LBP) e *Gradient Orientation Pyramid* (GOP) de cada rosto para explorar as informações discriminativas. Então, foi desenvolvido um método de aprendizagem discriminativa com representações de características *multiview*, chamado *Discriminative Learning Method* (DLM), para projetar diferentes tipos de características locais em um subespaço discriminativo. Utilizaram os conjunto de dados MORPH e FG-NET [Panis et al. 2016] nos experimentos. Utilizando o conjunto de dados MORPH o método MDL teve uma taxa de acurácia de 65,2%, e usando o FG-NET ele obteve uma taxa de 91,8%. O método foi comparado com os descritores de características locais, sendo usados individualmente ou em conjunto, e o método proposto superou todos.

Ali et al. [Ali et al. 2015] propuseram um algoritmo *Phase Congruency* (PC) combinado com *Local Binary Pattern Variance* (LBPV) para acomodar diferentes grupos etários durante a execução do AIFR. Os resultados dos experimentos mostraram que o desempenho das características combinadas podem superar o desempenho dos conjunto de recursos individuais (PC e LBPV). Usaram o conjunto de dados FG-NET para os experi-

mentos e obtiveram como resultado uma taxa de acurácia de 94,0%, superando consideravelmente os métodos propostos em [Li et al. 2011], [Park et al. 2010] e outros métodos.

Chen et al. [Chen et al. 2015] aproveitaram imagens de celebridades disponíveis na internet, para apresentar uma nova abordagem diferente dos estudos anteriores sobre AIFR. Em vez de modelar o processo de envelhecimento com suposições de parâmetros, usaram um método de codificação chamado *Cross-Age Reference Coding* (CARC). A premissa é que se duas pessoas são parecidas quando são jovens, elas também podem ser semelhantes quando envelhecem. Com base nisso, CARC utiliza um conjunto de imagens de referência disponíveis livremente na internet para codificar as características de baixo nível da imagem de uma face com uma representação de referência. Nos experimentos, foi utilizado o conjunto de dados MORPH para mostrar a eficácia do método, utilizando a mesma configuração de experimentos realizados em [Li et al. 2011]. Eles obtiveram uma taxa de acurácia de 92,8%, superando os métodos realizados em [Gong et al. 2013], [Li et al. 2011] e [Park et al. 2010]. Também desenvolveram seu próprio conjunto de dados chamado CACD-VS, obtendo uma taxa de acurácia de 87,6%.

Hou et al. [Hou et al. 2016] se concentraram, principalmente, em encontrar uma característica expressiva invariante na idade de tal forma que ela seja robusta para variações intrapessoais e discriminatória para indivíduos diferentes. Para isso, eles mapearam uma característica original para um novo espaço, no qual a característica é robusta a ruídos e a grande variações intrapessoais causadas por imagens faciais envelhecidas. Em seguida, codificaram ainda mais a característica mapeada em uma representação invariante na idade. Depois de mapear e codificar, obtiveram a característica robusta e discriminativa para o propósito específico de AIFR. Eles realizaram experimentos nos conjuntos de dados CACD e MORPH, usando o protocolo de Chen et al. [Chen et al. 2015]. Usando o CACD para avaliar o desempenho do método, compararam com os métodos CARC, HFA e HD-LBP [Barkan et al. 2013]. O método obteve resultados melhores do que os outros métodos em varias faixa etárias diferentes. Com o conjunto de dados MORPH, compararam o método com os métodos de [Gong et al. 2015], HFA, Li et al. [Li et al. 2015], CARC, HFA, Li et al. [Li et al. 2011] e Park et al. [Park et al. 2010]. O método empatou com o método combinado de [Gong et al. 2015], com uma taxa de acurácia de 94,5%, porém superou todos os outros métodos citados.

Essa [Essa 2016] propôs um descritor de características invariante na idade, transformando imagens faciais tomadas em uma idade diferente para faces uniformemente aleatórias esparsas, por meio de redes neurais de auto-codificação. Ele construiu um conjunto de referência para codificadores aleatórios esparsos, em que cada um é treinado com imagens em uma idade diferente para a mesma pessoa, com faces aleatórias e com os valores alvo para aprender as características de representação de alto nível. Usando faces aleatórias, a representação adquirida destaca características precisas que são invariantes ao longo do tempo. O método proposto foi avaliado no conjunto de dados CACD. O conjunto de teste foi dividido em três grupos de faixa etária diferentes: 2004 – 2006, 2007 – 2009 e 2010 – 2012, e comparado com quatro métodos diferentes: CARC, HP-LBP, classificador Simile e HFA. Em todos os grupos, o método proposto tem um desempenho melhor que os outros, com uma acurácia de 53,4%, 56,3% e 61,5% respectivamente.

Somada et al. [Somada et al. 2017] desenvolveram um método que emprega a verificação da segmentação de imagens de face, que consiste nas três etapas a seguir: (1) a

segmentação da imagem facial gera três sub-imagens regionais a partir da imagem da face de entrada; (2) uma pontuação correspondente é calculada usando características gradiente de um par, que consiste na imagem de entrada e uma imagem registrada para cada uma das três sub-imagens geradas e a imagem original, obtendo assim quatro pontuações correspondentes; (3) o classificador de verificação avalia o vetor de pontuação correspondente, e prevê a probabilidade posterior de que duas imagens correspondentes pertençam à mesma pessoa. Utilizaram o conjunto de dados FG-NET para esclarecer a eficácia do método. Adotaram o esquema *leave-one-out* seguindo a pesquisa de [Li et al. 2011]. O método teve uma taxa de acurácia 86,6%, superando os resultados das seguintes pesquisas: [Park et al. 2010], [Li et al. 2011], [Gong et al. 2013] e [Gong et al. 2015], porém teve uma taxa menor que a pesquisa de [Wen et al. 2016].

## 2.2. Abordagens CNNs (ACNNs)

Wen et al. [Wen et al. 2016] propuseram uma estrutura de AIFR para aprender as características de diferentes idades usando um modelo de Rede Neural Convolutiva (*Convolution Neural Network* – CNN). Em vez de aplicar diretamente o modelo de aprendizagem profunda, eles propuseram um novo modelo chamado CNN guiada por Fator Latente (*Latent Factor guided Convolutional Neural Network* – LF-CNN) para abordar especificamente a tarefa de AIFR, por meio da aprendizagem acoplada dos parâmetros de CNNs e da Análise de Identidade Latente (*Latent Identity Analysis* – LIA), as características faciais invariante na idade podem ser extraídas, sendo mais robustas às variações causadas pelo processo de envelhecimento ao longo do tempo. Eles realizaram vários experimentos para demonstrar a eficiência da proposta das LF-CNNs em vários conjuntos de dados incluindo MORPH Álbum 2, FG-NET, CACD-VS [Chen et al. 2015] e FLW. Os experimentos realizados com o conjunto de dados MORPH seguiram o mesmo esquema de divisão de [Gong et al. 2013]. Eles obtiveram como resultado que a LF-CNN, que teve uma taxa de acurácia de 97,51%, superou as taxas de reconhecimento de HFA, CARC, MEFA, MEFA + SILF + MLBP. Com o conjunto de dados FG-NET, usaram o mesmo regime de teste de [Li et al. 2011] e a abordagem LF-CNN teve uma taxa de acurácia de 88,1%, superando [Li et al. 2011], HFA e MEFA. Usando o CACD, eles também seguiram o mesmo regime de testes de [Li et al. 2011]. A abordagem LF-CNN obteve uma taxa de acurácia de 98,5%, superando HFA e CARC. Por fim, realizaram os testes com o conjunto de dados LFW e obtiveram uma taxa de 99,10% para abordagem LF-CNN.

Zheng et al. [Zheng et al. 2017] desenvolveram uma rede de reconhecimento profunda chamada CNN Guiada por Estimativa de Idade (*Age Estimation Guided Convolutional Neural Network* – AE-CNN) para separar as variações causadas pelo envelhecimento das características específicas que são estáveis. O modelo projetado pode aprender características invariante na idade para reconhecimentos de faces. Foram realizados experimentos com os bancos de dados MORPH Álbum 2 e CACD. Usando o banco de dados CACD, seguiram a mesma configuração dos experimentos de [Chen et al. 2014]. O método AE-CNN superou todos os métodos citados a seguir: HFA, CARC, GSM e CAN. Os experimentos realizados com o conjunto de dados MORPH, foram realizados com as mesmas configurações de experimentos de [Chen et al. 2014]. Usando o modelo de treinamento de CASIA-WebFace, o método AE-CNN teve uma taxa de reconhecimento e superou todos os seguintes métodos que ele foi comparado: LF-CNN, GSM, MEFA + SILF + MLBP, MEFA, CARC e HFA.

Yu e Jing [Yu and Jing 2018] propuseram uma estrutura de CNN Conjunta Multitarefa ou (*Joint Multi-task Convolutional Neural Network – JMCNN*). O JMCNN consiste em duas tarefas: uma para o reconhecimento facial aprendendo características invariantes na idade, e outra para classificação etária aprendendo características sensíveis à idade. Enquanto as duas tarefas se aprimoram, reforçam um termo de reconhecimento de regularização em dois tipos de características. Os experimentos foram realizados em dois bancos de dados MORPH Album 2 e CACD. O conjunto de dados CASIA-WebFace [Yi et al. 2014] é usado para treinar o modelo CNN. Usando o conjunto de dados: MORPH, eles seguiram o mesmo esquema de testes de [Babu and Kunju 2017], o método teve uma taxa de 98,36% no reconhecimento, superando todos os outros métodos que ele foi comparado, sendo eles: AE-CNN, LF-CNN, GMS, MEFA + SIFT + MLBP, MEFA, CARC e HFA. Usando o conjunto de dados CACD, seguiram as mesmas configurações de experimentos de [Chen et al. 2014], compararam o JMCNN com os métodos: CARC, CAN e AE-CNN, e o JMCNN superou todos os métodos citados.

### **2.3. Abordagem Híbrida (AH)**

Zhao et al. [Zhao et al. 2018] propuseram um Modelo Invariante na Idade Profundo ou (*Age-Invariant Model – AIM*) que consiste em uma Rede Generativa Adversária (*Generative adversarial network – GAN*), que contém uma sub-rede de Aprendizagem de Característica (*Representation Learning sub-Net – RLN*) e uma sub-rede de Síntese Facial (*Face Synthesis sub-Net – FSN*) que, em conjunto, aprendem características faciais discriminativas e que realizam um envelhecimento e rejuvenescimento facial baseado nas características aprendidas. Eles também criaram um novo conjunto de dados em grande escala *CrossAge Face Recognition – CAFR*, que contém 1.446.500 imagens de faces de 25.000 indivíduos descritos com a idade, nome, gênero, raça e uma descrição de referência. Eles realizaram os experimentos nos três bancos de dados MORPH Álbum2, CACD-VS, FG-NET e em seu próprio conjunto de dados CARF. Os experimentos realizados no conjunto de dados MORPH Álbum2 retornaram uma taxa de acurácia de até 99,5% superando todos os outros métodos citados acima. Utilizando o CACD-VS obteve uma taxa na verificação de até 99,76% também superando os outros métodos e com o conjunto de dados FG-NET retornou uma taxa de acurácia de 93,2%.

Em contrapartida, a abordagem proposta neste trabalho utiliza VGG-Face2 [Cao et al. 2018], com a arquitetura Se-Net-50, como extrator de características, seguido do classificador *Linear Support Vector Machines (Linear SVM)* [Joachims 2006]. Tal abordagem se mostrou comparável ao estado da arte, sendo superior em alguns testes.

## **3. Materiais e Métodos**

Nesta seção serão apresentadas os conjuntos de dados mais utilizadas pelos artigos revisados e a metodologia experimental empregadas com eles. Além disso, será explicado brevemente o modelo VGGFace2 e a metodologia adotada para o uso desse modelo.

### **3.1. Conjuntos de Dados**

Esta seção contém breves descrições dos conjuntos de dados mais utilizadas, relatados pela revisão da literatura deste trabalho, dentre os quais estão presentes o FG-NET e o CACD. O conjunto de dados MORPH não foi utilizada, pois ela não é um conjunto de dados gratuito. Também são descritas os conjuntos de dados VGGFace e VGGFace2 que são usadas para treinar modelos de CNNs profundas.

### 3.1.1. FG-NET

O FG-NET [Panis et al. 2016] é um conjunto de dados de imagens publicamente disponível, contendo imagens faciais que mostram diferentes indivíduos em diferentes idades. O conjunto de dados foi desenvolvido na tentativa de auxiliar pesquisadores que investigam os efeitos do envelhecimento na aparência facial. O conjunto de dados contém 1002 imagens de 82 indivíduos diferentes com idades variando entre recém-nascidos até 69 anos de idade. No entanto, idades entre zero e 40 anos são a maioria no conjunto de dados, sendo que 61% desse conjunto é composto de imagens de pessoas do sexo masculino e 39% de imagens de pessoas do sexo feminino.

### 3.1.2. CACD

Chen et al. [Chen et al. 2015] fizeram seu próprio conjunto de dados recolhendo imagens de celebridades pela internet. O *Cross-Idade Celebridade Dataset* (CACD) contém mais de 160.000 imagens de rosto de 2.000 celebridades em dez anos, com a idade variando de 16 à 62 anos. Cada celebridade é classificada com um *rank* que corresponde a classificação da celebridade, com o mesmo ano de nascimento no *IMDB.com* de quando o conjunto de dados foi construído. O CACD tem números aproximadamente iguais de indivíduos do sexo masculino e feminino, e contém principalmente pessoas brancas. Desenvolveram ainda um subconjunto do CACD, chamado CACD-VS, para avaliar o desempenho na verificação facial. O CACD-VS contém 2.000 pares positivos (imagens de uma mesma pessoa através da idade) e 2.000 pares negativos (imagens de pessoas diferentes).

### 3.1.3. Modelo VGGFace e VGGFace2

VGGFace e VGGFace2 são conjuntos de dados desenvolvidos para treinar modelos de CNNs profundas para o reconhecimento facial. No entanto, tais termos são usados para se referir aos modelos CNNs pré-treinados por esses conjuntos de dados.

O VGGFace2 [Cao et al. 2018] é definido como um conjunto de dados para reconhecer faces por meio de pose e idade. Ele contém 3,31 milhões de imagens de 9.131 indivíduos, com uma média de 362,6 imagens para cada indivíduo. As imagens são baixadas de pesquisas de imagens do Google e apresentam grandes variações de pose, idade, iluminação, etnia e profissão (por exemplo, atores, atletas, políticos). O modelo foi treinado utilizando duas arquiteturas de CNNs diferentes, a ResNet-50 e a SeNet-50. Comparando os modelos treinados, o SeNet-50 apresenta um desempenho superior no reconhecimento facial, que também pode ser utilizado como extrator de características.

## 3.2. Abordagem Proposta

Na revisão da literatura foi possível observar que os trabalhos mais recentes, com os melhores resultados para o problema proposto, utilizaram CNNs. Por isso, este trabalho também propôs o uso de CNNs. Entretanto, ao contrário da revisão de literatura, neste trabalho foi utilizada a CNN VGGFace2, com a arquitetura Se-Net-50, como extrator de características e diferentes métodos para a classificação, como *K-Nearest Neighbors* (K-NN), K-NN com distância do cosseno [Pirlo and Impedovo 2013], *Linear Discriminant*

*Analysis* (LDA) [Sungatullina et al. 2013] e *Linear SVM* [Joachims 2006]. Tais abordagens foram comparadas entre si e com as principais abordagens da revisão de literatura. Baseando-se nos resultados dos testes realizados neste trabalho, verificou-se que a melhor abordagem proposta por este trabalho foi a extração de características com o VGGFace2, seguido da classificação com o *Linear SVM*.

### **3.3. Metodologia Experimental**

Para realizar a comparação entre a abordagem proposta e os trabalhos da literatura, foram utilizados os mesmos métodos de validação cruzada utilizados pelos trabalhos revisados. A seguir são descritas as metodologias utilizadas com cada conjunto de dados.

#### **3.3.1. Experimentos com o Modelo VGGFace2 no conjunto de dados FG-NET**

Para iniciar os experimentos foi utilizado o conjunto de dados FG-NET, por ser um conjunto de dados gratuito e muito utilizado em outros experimentos com a abordagem *leave-one-out* [Kohavi 1995], também utilizada por outros autores para o cálculo da acurácia utilizando o conjunto de dados FG-NET. Assim, é possível comparar os resultados obtidos nesta pesquisa com o resultado de outros autores, sem a necessidade de implementar as abordagens propostas por eles. Nesse esquema é escolhida uma imagem como amostra de teste e o restante das imagens são utilizadas para o treinamento do modelo, isso é repetido para cada imagem do conjunto de dados, e no final é calculada a taxa de acurácia, dividindo o quantidade de acertos de cada teste pelo tamanho do conjunto de dados.

Utilizando o modelo VGGFace2 os descritores de cada imagem do conjunto de dados são extraídos e salvos em vetor de 1002 posições. Em seguida, uma função percorre o vetor e retira um descritor por vez para ser usado como descritor de teste. Os 1001 descritores restantes são usados para o treinamento do modelo. Por fim, o descritor extraído é enviado para o modelo de classificação, que retorna uma classe para ser comparada com a classe original do descritor, retornando se a classificação foi correta.

#### **3.3.2. Experimentos com o Modelo VGGFace2 na conjunto de dados CACD**

Para realizar os testes com o CACD, foi seguido a mesma configuração de divisão do conjunto de dados definida em Chen et al. [Chen et al. 2015]. Tal configuração foi utilizada por todos os trabalhos comparados. Foram usadas 120 celebridades que fazem parte dos *rank* de 3 a 5, classificadas pelos autores do conjunto de dados. As imagens foram divididas em três grupos de imagens diferentes, sendo cada grupo imagens tiradas entre 2004 – 2006, 2007 – 2009 e 2010 – 2012, que são as imagens usadas como conjunto de treino. Um quarto grupo de imagens, tiradas em 2013, é usado como o conjunto de teste. Em seguida, é extraído o descritor de cada imagem do conjunto de treino e teste, utilizando o modelo VGGFace2. Assim, é realizado o treinamento do classificador *Linear SVM* com cada uma das três bases de treino (2004 – 2006, 2007 – 2009 e 2010 – 2012). Finalmente, o classificador é testado usando a base de teste dos descritores das imagens de 2013, retornando três taxas de acurácia diferentes, uma para cada conjunto de treino. As taxas de acurácia são calculadas dividindo a quantidade de imagens classificadas corretamente, para cada treinamento realizado, pela quantidade de imagens do conjunto de teste.

#### 4. Resultados e Discussão

No conjunto de dados FG-NET foram realizados testes com os descritores retirados de duas CNNs pré-treinadas, VGGFace e VGGFace2, utilizando a arquitetura Se-Net-50. Nos primeiros testes, utilizando como extrator de características a CNN treinada com o conjunto de dados VGGFace2, obteve resultados melhores que os com VGGFace. Isso se deve pelo fato da VGGFace2 já ser um conjunto de dados mais focado em imagens com faces em diferentes poses e idades. Para a classificação dos descritores, foram testados vários métodos de classificação como o K-NN, LDA, K-NN com distância do cosseno. No entanto, o classificador que obteve melhor resultado foi o *Linear SVM* (veja a Tabela 1).

Classificador	Acurácia
K-NN com $K = 1$	84,4%
K-NN com distância do cosseno	85,0%
LDA	88,1%
<i>Linear SVM</i>	<b>89,9%</b>

Tabela 1. Testes com diferentes classificadores no conjunto de dados FG-NET.

Com a utilização do FG-NET, também foi possível observar que a abordagem tem mais dificuldade em reconhecer as faces quando os indivíduos têm idades mais jovens. Isso ocorre pelo fato da mudança facial de um indivíduo ser maior quando ele tem menos de 21 anos de idade, o que dificulta o trabalho do reconhecimento facial (veja a Figura 1).

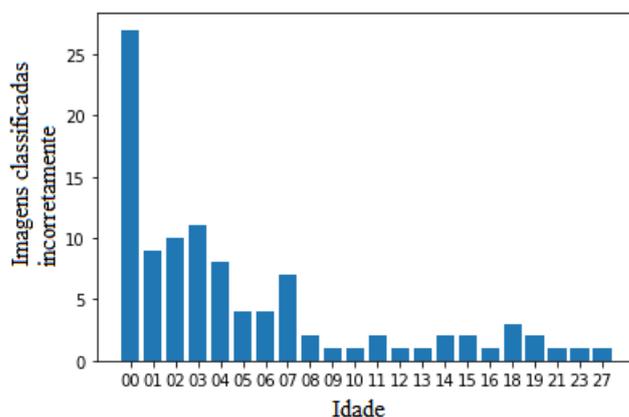


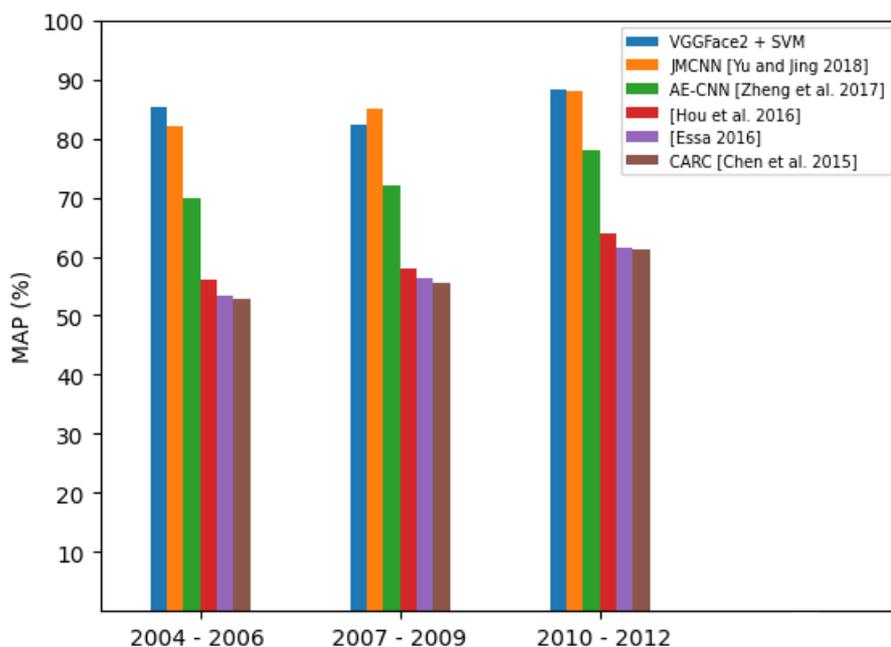
Figura 1. Gráfico de quantidade de erros por idade.

De acordo com a revisão da literatura, é possível observar que as abordagens utilizando CNN obtiveram resultados superiores em relação às outras. O experimento utilizando VGGFace2 e Linear SVM obtiveram uma taxa de acurácia de 89,9% sem a necessidade de nenhum pré-processamento das imagens de entrada. A Tabela 4 apresenta o resultado da acurácia de todas as abordagens citadas neste trabalho, que utilizaram o conjunto de dados FG-NET. Ela mostra que a abordagem proposta só está com um percentual de 3,4% abaixo do melhor resultado da Tabela 4.

Trabalho	Abordagem	Acurácia (%)
[Park et al. 2010]	Age-Invariant Face Recognition	37,40
[Li et al. 2011]	A Discriminative Model for Age Invariant Face Recognition (MFDA)	47,50
[Tandon et al. 2014]	An Efficient Age-Invariant Face Recognition	48,40
[Bouchaffra 2015]	Nonlinear Topological Component Analysis: Application to Age-Invariant Face Recognition (NTCA)	48,96
[Yang et al. 2014]	Multiview Discriminative Learning for Age-Invariant Face Recognition (TED-GM)	64,47
[Gong et al. 2013]	Hidden Factor Analysis for Age Invariant Face Recognition (HFA)	69,00
[Gong et al. 2015]	A Maximum Entropy Feature Descriptor for Age Invariant Face Recognition (MEFD)	76,20
[Alvi and Pears 2015]	Use of Spatio-Temporal Modeling for Age Invariant Face Recognition	85,04
[Somada et al. 2017]	Segmented Face Image Verification for Age-Invariant Face Recognition	86,60
[Wen et al. 2016]	Latent Factor Guided Convolutional Neural Networks for Age-Invariant Face Recognition (LF-CNN)	88,10
[Sungatullina et al. 2013]	Multiview Discriminative Learning for Age-Invariant Face Recognition (MDL)	91,80
[Zhao et al. 2018]	Look Across Elapse: Disentangled Representation Learning and Photorealistic Cross-Age Face Synthesis for Age-Invariant Face Recognition	93,20
[Ali et al. 2015]	Age-invariant face recognition system using combined shape and texture features	93,33
Este trabalho	VGGFace2 com Linear SVM	<b>89,90</b>

**Tabela 2. Resultados utilizando FG-NET.**

Os experimentos realizados com o conjunto de dados CACD também retornaram bons resultados. Para cada conjunto de treino, é calculado três taxas de acurácia diferentes, 85,2%, 82,4% e 88,2%, respectivamente, como pode ser visto na Figura 2, em que a abordagem proposta supera o JMCNN [Yu and Jing 2018] com a bases de treino de 2004 – 2006 com 82% e 2010 – 2012 com 85%, e todos os outros trabalhos citados.



**Figura 2. Resultados utilizando o conjunto de dados CACD.**

A vantagem de se utilizar VGGFace2 em relação às outras abordagens propostas,

é que ela já é uma rede pré-treinada. Assim, não é necessário treinar um novo modelo de CNN, que requer um grande custo computacional e um longo período de tempo.

## 5. Conclusão e Trabalhos Futuros

Neste trabalho foi proposta a utilização de uma CNN profunda pré-treinada para a classificação de imagens de faces em diferentes idades. De acordo com os resultados obtidos nos experimentos, é possível observar que a abordagem proposta utilizando VGGFace2 e Linear SVM é bem eficaz para ser usada no problema de AIFR. Comparado-a com outras abordagens, o método de extrair os descritores da imagem acaba obtendo um resultado mais rápido e com menos custo computacional.

Em trabalhos futuros serão desenvolvidos e realizados testes acessando a abordagem proposta por meio de um dispositivo *wearable*, para que seja possível verificar sua viabilidade no auxílio do reconhecimento facial para pessoas com deficiência visual em relação ao problema do envelhecimento. Além disso, serão realizadas análises sobre a economia dos recursos de hardware do dispositivo durante o uso da abordagem proposta.

## Referências

- Ali, A. S. O., Sagayan, V., Saeed, A. M., Ameen, H., and Aziz, A. (2015). Age-invariant face recognition system using combined shape and texture features. *IET Biometrics*, 4(2):98–115.
- Babu, K. and Kunju, N. (2017). Cross age face recognition using deep learning model based on center symmetric lps. *ICECDS*, pages 1897–1900.
- Barkan, O., Weill, J., Wolf, L., and Aronowitz, H. (2013). Fast high dimensional vector multiplication face recognition. In *IEEE ICCV*, pages 1960–1967.
- Britto Neto, L., Grijalva, F., Maïke, V. R. M. L., Martini, L. C., Florencio, D., Baranauskas, M. C. C., Rocha, A., and Goldenstein, S. (2017). A kinect-based wearable face recognition system to aid visually impaired users. *IEEE THMS*, 47(1):52–64.
- Britto Neto, L., Maïke, V. R. M. L., Koch, F. L., Baranauskas, M. C. C., Rocha, A., and Goldenstein, S. (2015). A wearable face recognition system built into a smartwatch and the visually impaired user. In *ICEIS*, pages 5–12. INSTICC, SciTePress.
- Britto Neto, L. S. (2016). *Sistemas Wearable baseados em Métodos de Visão Computacional para Auxiliar Pessoas com Deficiência Visual*. PhD thesis, UNICAMP, Campinas.
- Cao, Q., Shen, L., Xie, W., Parkhi, O. M., and Zisserman, A. (2018). Vggface2: A dataset for recognising faces across pose and age. In *IEEE FG*, pages 67–74.
- Chen, B., Chen, C., and Hsu, W. H. (2015). Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Trans. Multi.*, 17(6):804–815.
- Chen, B.-C., Chen, C.-S., and Hsu, W. H. (2014). Cross-age reference coding for age-invariant face recognition and retrieval. *ECCV*, pages 768–783.
- Essa, E. (2016). Sparse random encoder for age invariant face recognition. *2016 11th International Conference on Computer Engineering Systems (ICCES)*, pages 167–171.
- Gong, D., Li, Z., Lin, D., Liu, J., and Tang, X. (2013). Hidden factor analysis for age invariant face recognition. *2013 IEEE ICCV*, pages 2872–2879.

- Gong, D., Li, Z., Tao, D., Liu, J., and Li, X. (2015). A maximum entropy feature descriptor for age invariant face recognition. *IEEE CVPR*, pages 5289–5297.
- Hou, X., Ding, S., and Ma, L. (2016). Robust feature encoding for age-invariant face recognition. *IEEE ICME*, pages 1–6.
- Joachims, T. (2006). Training linear svms in linear time. In *ACM SIGKDD, KDD '06*, pages 217–226, New York, NY, USA. ACM.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI*, pages 1137–1143.
- Li, Y., Wang, G., Lin, L., and Chang, H. (2015). A deep joint learning approach for age invariant face verification. In Zha, H., Chen, X., Wang, L., and Miao, Q., editors, *Computer Vision*, pages 296–305, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Li, Z., Park, U., and Jain, A. K. (2011). A discriminative model for age invariant face recognition. *IEEE TIFS*, 6(3):1028–1037.
- Panis, G., Lanitis, A., Tsapatsoulis, N., and Cootes, T. F. (2016). Overview of research on facial ageing using the fg-net ageing database. *IET Biometrics*, 5(2):37–46.
- Park, U., Tong, Y., and Jain, A. K. (2010). Age-invariant face recognition. *IEEE TPAMI*, 32(5):947–954.
- Pirlo, G. and Impedovo, D. (2013). Cosine similarity for analysis and verification of static signatures. *IET Biometrics*, 2(4):151–158.
- Ricanek, K. and Tesafaye, T. (2006). Morph: a longitudinal image database of normal adult age-progression. *FGR*, pages 341–345.
- Somada, Y., Ohshima, W., and Wakabayashi, T. (2017). Segmented face image verification for age-invariant face recognition. *ICIEV-ISCMT*, pages 1–4.
- Sungatullina, D., Lu, J., Wang, G., and Moulin, P. (2013). Multiview discriminative learning for age-invariant face recognition. *IEEE FG*, pages 1–6.
- Wen, Y., Li, Z., and Qiao, Y. (2016). Latent factor guided convolutional neural networks for age-invariant face recognition. *IEEE CVPR*, pages 4893–4901.
- Yang, H., Huang, D., and Wang, Y. (2014). Age invariant face recognition based on texture embedded discriminative graph model. *IEEE IJCB*, pages 1–8.
- Yi, D., Lei, Z., Liao, S., and Li, S. Z. (2014). Learning face representation from scratch. *CoRR*, abs/1411.7923.
- Yu, J. and Jing, L. (2018). A joint multi-task cnn for cross-age face recognition. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 2411–2415.
- Zhao, J., Cheng, Y., Cheng, Y., Yang, Y., Lan, H., Zhao, F., Xiong, L., Xu, Y., Li, J., Pranata, S., Shen, S., Xing, J., Liu, H., Yan, S., and Feng, J. (2018). Look across elapse: Disentangled representation learning and photorealistic cross-age face synthesis for age-invariant face recognition. *CoRR*, abs/1809.00338.
- Zheng, T., Deng, W., and Hu, J. (2017). Age estimation guided convolutional neural network for age-invariant face recognition. *IEEE CVPRW*, pages 503–511.