Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Reconhecimento automático de traços de personalidade em vídeo

André Barbosa Carneiro da Cunha Bauer

Monografia apresentada como requisito parcial
para conclusão do Curso de Engenharia da Computação

Orientador
Prof. Dr. Flávio de Barros Vidal

Brasília
2017

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Reconhecimento automático de traços de personalidade em vídeo

André Barbosa Carneiro da Cunha Bauer

Monografia apresentada como requisito parcial
para conclusão do Curso de Engenharia da Computação

Prof. Dr. Flávio de Barros Vidal (Orientador)
CIC/UnB

Prof. Dr. Bruno Macchiavello
CIC/UnB

Prof. Dr. Teófilo Emidio De Campos
CIC/UnB

Prof. Dr. Ricardo Pezzuol Jacobi
Coordenador do Curso de Engenharia da Computação

Brasília, 17 de abril de 2017

Dedicatória

Dedico esse trabalho a toda minha família, principalmente minha mãe Isabel Botelho Barbosa, meu padrasto Pedro Luiz Dalcero, meu irmão Pedro Barbosa Carneiro da Cunha Bauer e minha avó Jandacy Leal Barbosa pelo apoio e tudo que fizeram por mim.

Agradecimentos

Agradeço ao meu orientador, Prof. Dr. Flávio de Barros Vidal pela paciência, profissionalismo e compreensão. À minha família por todo o apoio. Aos meus colegas e amigos de curso: Ana Paula Almeida, Tiago Pigatto Lenza, Victor Fernandes Uriarte por todos os obstáculos e caminhos atravessados juntos, assim como pela companhia.

Resumo

A análise de traços de personalidade é uma área em crescimento tanto na psicologia, como na computação, devido à sua grande utilização como indicador de resultados importantes da vida, como felicidade e longevidade, qualidade do relacionamento com colegas, família, escolha de ocupação, satisfação e desempenho, envolvimento na comunidade, atividade criminal e ideologia política.

Neste trabalho é proposto um método de análise de traços de personalidade a partir de informações visuais. Isto é feito a partir de uma base de dados de vídeos, de aproximadamente 15 segundos, de indivíduos sozinhos em primeiro plano em frente a uma câmera. A implementação foi realizada por meio da combinação de técnicas de Fluxo Ótico e da obtenção de Histogramas de Gradientes Orientados, aplicados em uma *perceptron* multicamadas, modelo de Rede Neural .

A métrica de avaliação utilizada foi a da acurácia média, que representa a distância média do resultado obtido ao resultado esperado.

Para trabalhos futuros, serão definidas novas características para a rede neural, assim como será utilizado o áudio dos vídeos, podendo assim a implementação ser adaptada para um detector de mentiras.

Palavras-chave: redes neurais, fluxo ótico, histograma de gradientes orientados

Abstract

Personality traits analysis is a field with growing interest not only in psychology, but also in computing, due to its importance as a predictor of important life outcomes, such as happiness, longevity, quality of relationships with peers, family, occupational choice, satisfaction, performance, community involvement, criminal activity and political ideology.

This project proposes a method for personality trait analysis from visual information. This work is based on a database that contains videos of approximately 15 seconds of individuals on the foreground in front of a camera. The implementation was performed by using the combination of Optical Flow techniques and Histograms of Oriented Gradients, applied to a multi-layer perceptron model of a neural network.

A mean accuracy metric was used for evaluation, which is denoted by the average distance from the obtained result to the expected result.

For future works, new characteristics will be specified for the neural network and the audio will also be used, so that the implementation can be adapted for lie detection.

Keywords: neural network, optical flow, histogram of oriented gradients

Sumário

1 Introdução **1**
1.1 Objetivos . . . 1
1.1.1 Principal . . . 1
1.1.2 Secundários . . . 1
1.2 Motivação . . . 1
1.3 Trabalhos Relacionados . . . 2
1.4 Divisão do Trabalho . . . 4

2 Fundamentação Teórica **5**
2.1 Sequências de Imagens . . . 5
2.2 Fluxo Ótico . . . 5
2.3 Histograma de Gradientes Orientados . . . 8
2.4 Rede Neural Artificial . . . 10

3 Metodologia **13**
3.1 Base de dados . . . 13
3.1.1 Big five . . . 14
3.1.2 Fluxo Ótico . . . 16
3.1.3 Histograma de Gradientes Orientados . . . 17
3.1.4 Rede Neural Artificial . . . 19
3.1.5 Métrica de Avaliação . . . 23

4 Resultados **24**
4.1 Análise dos Resultados . . . 24
4.1.1 Cenário 1 . . . 26
4.1.2 Cenário 2 . . . 30
4.1.3 Resultados Finais . . . 31

5 Conclusão e Tabalhos Futuros **33**

Referências **34**

Lista de Figuras

2.1 Aplicação de um descritor HOG de células de tamanho diferente. Retirado de [1]. 9
2.2 Exemplo de uma rede perceptron multicamadas, adaptado de [2]. 11

3.1 Fluxograma da Metodologia Proposta. 13
3.2 Amostras da base de dados. 15
3.3 Resumo do *Big Five*, retirado de [3]. 16
3.4 Gráfico da média da acurácia variando número de camadas e modo de normalização do HOG. 18
3.5 Gráfico variando camadas normalizando HOG dividindo-o pela resolução (m*n). 20
3.6 Gráfico variando camadas normalizando HOG dividindo-o por m+n. 20
3.7 Gráfico variando camadas normalizando HOG dividindo-o por fr(m*n). . . 21
3.8 Gráfico variando camadas normalizando HOG dividindo-o por fr(m+n). . . 21
3.9 Gráfico variando número de neurônios. 22
3.10 Gráfico da média da acurácia variando número de neurônios entre 1 e 10. . 22

4.1 Gráfico da média da acurácia variando porcentagem utilizada para treinamento e para validação. 25
4.2 Gráfico do resultado de acurácia final. 25
4.3 Gráfico dos resultados da configuração inicial. Cada cor corresponde a um traço de personalidade. 26
4.4 Gráfico dos resultados do cenário 1. 28
4.5 Gráfico dos resultados do cenário 1. 29
4.6 Gráfico dos resultados do cenário 1. Cada cor corresponde a um traço de personalidade. 30
4.7 Gráfico dos resultados da configuração final. 31

Lista de Tabelas

4.1 Comparação do resultados deste trabalho com trabalho com os do desafio . 32

Lista de Abreviaturas e Siglas

AMT *Amazon Mechanical Turk*.

CNN Rede Neural Convolucional.

DAN *Descriptor Aggregation Networks*.

HOG Histograma de Gradientes Orientados.

LSTM Memória de Curto a Longo Prazo.

MCF Modelo de Cinco Fatores.

MLP *Perceptron* Multicamadas.

OpenCV *Open Computer Vision*.

PCA Análise de Componentes Principais.

RNA Rede Neural Artificial.

RNN Rede Neural Recorrente.

SSD Soma das Diferenças Quadráticas.

Capítulo 1

Introdução

1.1 Objetivos

1.1.1 Principal

Este trabalho visa desenvolver uma metodologia para extração automática de traços de personalidade de um indivíduo a partir de vídeos de uma base de dados.

1.1.2 Secundários

Para a extração automática de traços de personalidade, serão definidas características a serem derivadas das imagens dos vídeos para aplicá-las a um classificador supervisionado para a determinação automática desses traços.

1.2 Motivação

Atualmente, a maioria das grandes empresas possui psicólogos avaliando candidatos a vagas e geralmente a primeira impressão é a que mais marca. De acordo com [4], primeiras impressões são julgamentos de traços de personalidade e características sociais complexas como dominância, hierarquia, cordialidade e ameaça [5, 6, 7]. Foi demonstrado que traços de personalidade podem ser definidos de forma precisa quando os observadores são expostos a fluxos de imagens contínuos de comportamento de indivíduos [5, 8] por intervalos de tempo relativamente curtos (4 a 10 minutos) e até mesmo se expostos a imagens estáticas pelo curto período de 10s. A designação desses traços por observadores humanos podem ser realizadas tão rapidamente quanto 100ms [9].

A análise de personalidade caracteriza um bom indicador para resultados importantes da vida, como felicidade e longevidade, qualidade do relacionamento com colegas, família,

escolha de ocupação, satisfação e desempenho, envolvimento na comunidade, atividade criminal e ideologia política [10, 11].

A suposição fundamental da psicologia sobre personalidade é que características individuais resultam em padrões comportamentais estáveis que pessoas tendem a mostrar independentemente da situação [12]. O Modelo de Cinco Fatores (MCF) é o paradigma dominante em pesquisa de personalidade [4].

No campo da Ciência da Computação, a Computação de Personalidade estuda como máquinas poderiam reconhecer ou sintetizar automaticamente a personalidade humana [12]. Métodos desde aspectos não verbais da comunicação, como combinações multimodais de estilo de fala e movimentos corporais, expressões faciais, combinando acústica com sinais visuais ou fisiologia com esses sinais, foram propostos para reconhecimento de personalidade.

1.3 Trabalhos Relacionados

A classificação de traços de personalidade vem recebido bastante atenção por pesquisadores devido à sua relevância em estudos do comportamento humano e interface computador-humano [4].

Em [13], é apresentada uma abordagem em tempo real de reconhecimento de emoções. Primeiramente é usada uma detecção de face para em seguida, a partir de características obtidas baseadas na extração de propriedades da localização dos órgãos do indivíduo como olhos, boca e nariz, assim como a relação entre eles e de uma base de emoções como comparação; classificar a emoção presente no instante.

No artigo supracitado, utiliza-se uma *webcam* para capturar a imagem de entrada, quadro a quadro. Nessa imagem, é feito um preprocessamento como equalização de histogramas e detecção de bordas, para manter a eficiência do sistema. A partir da imagem obtida, a detecção de face é realizada por meio de detectores baseados em Haar [14].

É realizado um treinamento utilizando Eigenface [15], com fins de detecção da face humana, e Análise de Componentes Principais (*Principal Component Analysis* - PCA), classifica-se as emoções presentes nos quadros subsequentes.

A plataforma de código aberto *CodaLab* fornece um sistema para pesquisas computacionais visando um modo eficiente, reproduzível e colaborativo entre pesquisadores. Do desafio de [4] desta plataforma, em que se baseou este trabalho, pode-se extrair métodos com o mesmo objetivo do presente neste trabalho, alguns dos quais serão descritos brevemente a seguir. As três equipes mais bem classificadas abordaram o problema utilizando fluxos separados para áudio e vídeo, aplicando redes neurais para ambos. Os dois primeiros lugares realizaram algum tipo de preprocessamento, o primeiro usou características

logfbank, um filtro para o áudio e o segundo utilizou corte da face e características de áudio espectral.

A equipe *NJU-LAMDA*, que ficou em primeiro lugar, propôs dois modelos separados para imagens e áudio, processando múltiplos quadros do vídeo e empregando uma fusão tardia em duas etapas das previsões de quadro e áudio. Para o vídeo, foi proposta a utilização de um DAN+, extensão de *Descriptor Aggregation Networks* [16] aplica um agrupamento de máximo e média em duas camadas da Rede Neural Convolucional (*Convolutional Neural Network* - CNN) [17], normalizando e concatenando as saídas antes de alimentá-las às camadas totalmente conectadas. Um modelo pré-treinado VGG-face [18] é utilizado, substituindo as camadas totalmente conectadas e ajustando finamente o modelo com o conjunto de dados. Para o áudio são empregadas características de um banco de filtros de registro (*logfbank*) e uma camada totalmente conectada com ativações sigmoidais. Na fase de testes, um número pré-determinado de quadros é alimentado à rede visual e as predições arredondadas. Os resultados finais das predições são arredondados novamente com a saída do preditor de áudio.

A proposta da equipe *evolgen*, segundo colocado do desafio, foi de uma arquitetura multimodal de Memória de Curto a Longo Prazo (*Long Short-Term Memory* - LSTM), uma Rede Neural Recorrente (*Recurrent Neural Network* - RNN) [19], para predição dos traços de personalidade.

Para manter a estrutura temporal, as sequências de vídeo de entrada são separadas em seis partições sem sobreposição. Para cada partição, a representação do áudio é extraída usando características espectrais clássicas e medidas estatísticas, formando um vetor de características de 68 dimensões. A representação do vídeo é extraída selecionando aleatoriamente um quadro da partição, extraindo a face e a centrando utilizando alinhamento de face. O dado pré-processado é passado a uma CNN Recorrente, treinado ponta a ponta, usando um *pipeline* separado para áudio e vídeo. Cada partição de quadro é processado com camadas convolucionais, depois aplicado a uma transformada linear para reduzir a dimensionalidade.

As características de áudio de uma partição dada passam por uma transformada linear e são concatenadas com as características do quadro. A camada Recorrente é sequencialmente alimentada com as características extraídas de cada partição. Dessa forma, a rede recorrente captura variações de áudio e de expressões faciais para a predição de traços de personalidade.

A proposta de terceiro lugar, que foi feita pela equipe *DCC*, é de um modelo de reconhecimento de traços de personalidade multimodal composto por dois fluxos auditivos e visuais distintos (redes de aprendizagem residual profunda [20], de 17 camadas cada), seguidos por um fluxo audiovisual que é treinado de ponta a ponta para predição dos traços

de personalidade. Não possui pré-treinamento, porém um simples pré-processamento é executado onde são selecionados aleatoriamente um quadro e um recorte do áudio como entrada. Durante os testes, o áudio e vídeo inteiros são alimentados nos fluxos auditivos e visuais, aplicando um agrupamento médio antes de ser alimentado a camada totalmente conectada.

1.4 Divisão do Trabalho

A divisão deste projeto é feita da seguinte forma: o Capítulo 2 apresenta os principais conceitos teóricos, fundamentais para a produção do trabalho. A metodologia proposta se encontra no Capítulo 3, os resultados obtidos e discussões estão no Capítulo 4 e trabalhos futuros e conclusão estão descritos no Capítulo 5.

Capítulo 2

Fundamentação Teórica

Neste capítulo, serão apresentados trabalhos relacionados assim como as técnicas utilizadas para realização desse.

2.1 Sequências de Imagens

Uma imagem pode ser definida como uma função bidimensional, $f(x, y)$, onde x e y são coordenadas espaciais e a amplitude de f em qualquer par de coordenadas (x, y) é chamado de intensidade de níveis de cinza da imagem naquele ponto [21].

Pela definição de Trucco [22], uma sequência de imagens é uma série de N imagens, ou quadros, adquiridos em instantes de tempo discretos $t_k = t_0 + k\Delta t$, onde Δt é um período de amostragem fixo, e $k = 0, 1, ..., N - 1$.

2.2 Fluxo Ótico

O Fluxo Ótico é um método para a estimação do movimento do padrão de brilho da imagem, onde este movimento é descrito por um vetor representando este deslocamento [23].

O fluxo ótico foi desenvolvido devido a necessidade de medir a velocidade da sequência de imagens. Com isso, vários métodos apareceram e são utilizados. Devido ao número crescente de técnicas de Fluxo Ótico e pela falta de avaliação dos métodos existentes, [24] avaliou e comparou algumas dessas técnicas de forma empírica e concentrou-se na acurácia, confiabilidade e densidade das medidas de velocidade e concluiu que seu desempenho varia significativamente de acordo com as técnicas implementadas.

O Fluxo Ótico pode ser obtido por meio de três técnicas diferentes, sendo elas: métodos diferenciais, métodos de similaridade de região e métodos baseados em energia.

Técnicas diferenciais

A técnica empregada neste trabalho é a proposta por Horn-Schunck [23] como método diferencial, em que calcula-se a estimação da velocidade do campo de movimento da imagem a partir de derivadas espaço-temporais da intensidade do padrão de brilho. As primeiras instâncias usavam derivadas de primeira ordem e foram baseadas em translação de imagens, ou seja onde $\mathbf{v} = (u, v)^T$.

$$\mathbf{I}(\mathbf{x}, t) = \mathbf{I}(\mathbf{x} - \mathbf{v}t, 0), \tag{2.1}$$

De uma expansão de Taylor da Equação 2.1 ou mais geralmente de uma suposição de que a intensidade é conservada, $d\mathbf{I}(x, t)/dt = 0$, a equação de restrição de gradiente é facilmente derivada:

$$\nabla\mathbf{I}(\mathbf{x}, t) \cdot \mathbf{v} + \mathbf{I}_t(\mathbf{x}, t) = 0, \tag{2.2}$$

onde $\mathbf{I}_t(\mathbf{x}, t)$ denota a derivada parcial de tempo de $\mathbf{I}(\mathbf{x}, t)$, $\nabla\mathbf{I}(\mathbf{x}, t) = (\mathbf{I}_\mathbf{x}(\mathbf{x}, t), \mathbf{I}_y(\mathbf{x}, t))^T$, e $\nabla I \cdot v$ indica o produto escalar comum. Com efeito, a Equação 2.2 produz a componente normal de movimento de contornos espaciais de intensidade constante, $\mathbf{v_n} = sn$. A normal de velocidade s e a normal de direção n são dadas por

$$\mathbf{s}(\mathbf{x}, t) = \frac{-\mathbf{I}_t(\mathbf{x}, t)}{||\nabla\mathbf{I}(\mathbf{x}, t)||}, \mathbf{n}(\mathbf{x}, t) = \frac{\nabla\mathbf{I}(\mathbf{x}, t)}{||\nabla\mathbf{I}(\mathbf{x}, t)||}. \tag{2.3}$$

Existem duas componentes desconhecidas de $\mathbf{v}$ na Equação 2.2, restringidas por uma equação linear. Restrições adicionais são portanto necessárias para resolver para os dois componentes de $\mathbf{v}$.

Métodos de diferenciação de segunda ordem usam derivadas de segunda ordem (matriz Hessiana de $\mathbf{I}$) para restringir velocidade em duas dimensões:

$$\begin{bmatrix} \mathbf{I}_{xx}(\mathbf{x}, t) & I_{yx}(\mathbf{x}, t) \\ \mathbf{I}_{xy}(\mathbf{x}, t) & I_{yy}(\mathbf{x}, t) \end{bmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} \mathbf{I}_{tx}(\mathbf{x}, t) \\ \mathbf{I}_{ty}(\mathbf{x}, t) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2.4}$$

A Equação 2.4 pode ser derivada da Equação 2.1 ou da conservação de $\nabla\mathbf{I}(\mathbf{x}, t)$, $d\nabla\mathbf{I}(\mathbf{x}, t)/dt = 0$. Estritamente falando, a conservação de $\nabla\mathbf{I}(\mathbf{x}, t)$ implica que as deformações de primeira ordem da intensidade não deveriam estar presentes. Isto é então uma restrição mais forte que a Equação 2.2 para campos de movimento admissíveis. Para medir a velocidade na imagem, assumindo que $d\nabla\mathbf{I}(\mathbf{x}, t)/dt = 0$, as restrições da Equação 2.4 podem ser usadas isoladas ou juntamente com a Equação 2.2 para produzir um sistema superdeterminado de equações lineares.

No artigo supracitado são apresentadas quatro técnicas diferenciais, porém neste trabalho será apresentada somente a utilizada, a de Horn e Schunck.

Horn e Schunck [23] combinaram a restrição de gradiente da Equação 2.2 com um termo de suavidade global para restringir o campo de velocidade estimado $\mathbf{v}(\mathbf{x}, t) = (u(\mathbf{x}, t), v(\mathbf{x}, t))$, minimizando

$$\int_D (\nabla \mathbf{I} \cdot \mathbf{v} + \mathbf{I}_t)^2 + \lambda^2 (||\nabla u||_2^2 + ||\nabla v||_2^2) d\mathbf{x} \tag{2.5}$$

definida sobre um domínio D, onde a magnitude de λ reflete a influência do termo de suavidade. Equações iterativas são usadas para minimizar a Equação 2.5 e obter a velocidade da imagem:

$$\begin{aligned} u^{k+1} &= \bar{u}^k - \frac{I_x[I_x\bar{u}^k + I_y\bar{v}^k + I_t]}{\alpha^2 + I_x^2 + I_y^2} \\ v^{k+1} &= \bar{v}^k - \frac{I_y[I_x\bar{u}^k + I_y\bar{v}^k + I_t]}{\alpha^2 + I_x^2 + I_y^2}, \end{aligned} \tag{2.6}$$

onde k denota o número da iteração, u^0 e v^0 denotam a estimativa de velocidade inicial que estão definidos como zero, e $\bar{u}^k$ e $\bar{v}^k$ denotam as médias de tons de cinza de vizinhança de u^k e v^k.

Existem outro métodos de diferenciais importantes, como o de Lucas e Kanade [25], porém não serão discutidos neste trabalho.

Métodos de Similaridade de Região

A diferenciação numérica precisa pode ser impraticável devido ao ruído, devido a um número pequeno de quadros existentes ou devido ao serrilhamento no processo de aquisição de imagem. Nesses casos, uma abordagem diferencial pode ser inapropriada e é natural se voltar para similaridade baseada na região [26, 27, 28, 29, 30]. Essa abordagem define velocidade $\mathbf{v}$ como deslocamento $\mathbf{d} = (d_x, d_y)$ que produz o melhor ajuste entre regiões de imagens em tempos diferentes. Encontrando a melhor quantidade de correspondências que maximiza a medida de similaridade, como a correlação cruzada normalizada ou minimizando a medida de distância, como a soma das diferenças quadráticas (Sum-of-Squared Difference - SSD):

$$
\begin{aligned}
SSD_{1,2}(x;\mathbf{d}) &= \sum_{j=-n}^{n}\sum_{i=-n}^{n}\mathbf{W}(i,j)[\mathbf{I}_1(x+(i,j)) - \mathbf{I}_2(x+\mathbf{d}+(i,j))]^2 \\
&= \mathbf{W}(x) * [\mathbf{I}_1(x) - \mathbf{I}_2(x+\mathbf{d})]^2,
\end{aligned}
\tag{2.7}
$$

onde W denota uma função discreta bidimensional, e $\mathbf{d} = (d_x, d_y)$ assume valores inteiros.

Existe uma relação próxima entre as medidas de distância SSD, a medida de similaridade da correlação cruzada e técnicas diferenciais. Minimizar os valores de distância SSD para maximizar a integral do termo do produto $\mathbf{I}_1(x)\mathbf{I}_2(x+\mathbf{d})$. A diferença na Equação 2.7 pode também ser vista como a média de peso com janelamento da aproximação de primeira ordem da derivada temporal de $\mathbf{I}(x,t)$.

Existem alguns técnicas como a de Anandan [26] e a de Singh [31] para essa método de fluxo ótico, porém não serão abordados neste trabalho.

Métodos Baseados em Energia

Uma terceira classe de técnicas de fluxo ótico é baseada na energia de saída de filtros com ajuste de velocidade [32, 33, 34, 35, 36, 37]. Esses métodos, também chamados de métodos baseado em frequência, devido à concepção de filtros com ajuste de velocidade no domínio de Fourier [38, 39, 40, 41]. A transformada de Fourier de um padrão de translação bidimensional é dada por

$$
\hat{\mathbf{I}}(k,\omega) = \hat{\mathbf{I}}_0(k)\delta(\omega + v^T k), \tag{2.8}
$$

onde $\hat{\mathbf{I}}_0(\mathbf{k})$ é a transformada de Fourier de $\mathbf{I}(x,0)$, $\delta(k)$ é uma função de impulso, ω denota a frequência temporal e $\mathbf{k} = (k_x, k_y)$ denota a frequência espacial. Isso mostra que todas potências diferentes de zero associadas a padrões de translação bidimensionais se encontram em um plano que passa pela origem no domínio da frequência. Foi mostrado que alguns métodos baseados em energia são equivalentes a métodos baseados em correlação [38, 40] e a abordagens baseadas em gradiente de Lucas e Kanade [32, 42].

Técnicas como a de Heeger [36] são utilizadas para esse método, porém não serão aprofundadas nesse trabalho.

2.3 Histograma de Gradientes Orientados

O Histograma de Gradientes Orientados (HOG) é um descritor de características [43] utilizado em visão computacional e processamento de imagens focado em detecção de

objetos. O método consiste em dividir a imagem em pequenas regiões chamadas de células e contar a ocorrência da orientação de gradientes em porções da imagem.

Em [43], foram estudado os descritores baseados em borda e gradiente e foi demonstrado experimentalmente que grades de descritores de HOG superam significativamente os conjuntos de recursos existentes para detecção humana, como Haar *wavelets* [44] e contexto de forma [45].

Para a implementação do HOG, foram utilizados descritores remanescentes de histogramas de orientação de borda, descritores de transformada de característica invariante à escala (SIFT) [46] e contextos de forma [45], que foram computados em uma grade de células uniformemente espaçadas usando sobreposição de normalizações de contraste locais para melhora de desempenho.

O método é implementado dividindo a imagem em pequenas regiões espaciais denominadas células e para cada uma é acumulado um histograma de uma dimensão das direções de gradiente ou de orientação de borda. A combinação das entradas do histograma formam a representação. Para melhor invariância na iluminação ou sombreamento, também pode ser usado uma normalização de contraste do local antes do método.

Na Figura 2.1 é apresentado um exemplo da utilização do HOG, onde a primeira imagem é a imagem de teste, a segunda e a terceira são as imagens da computação do HOG.

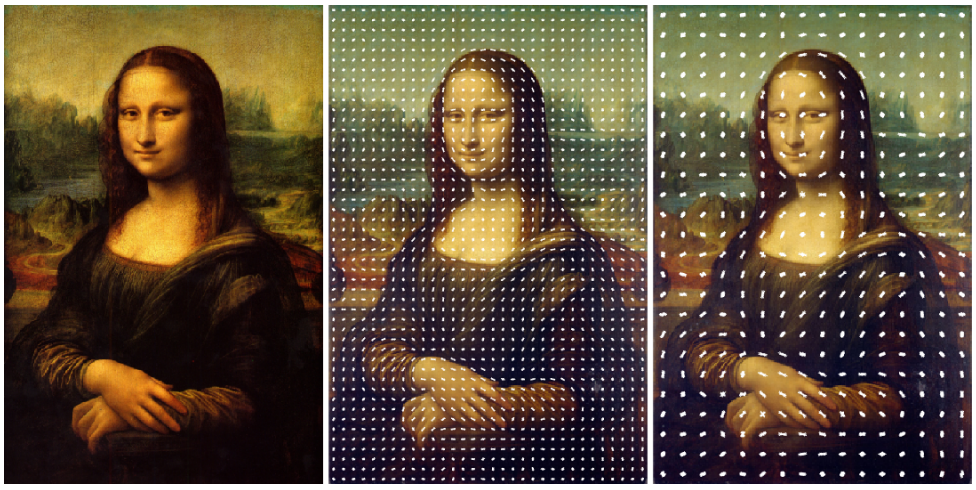

Figura 2.1: Aplicação de um descritor HOG de células de tamanho diferente. Retirado de [1].

2.4 Rede Neural Artificial

Segundo [2], Redes Neurais Artificiais (RNA) são modelos computacionais inspirados no sistema nervoso humano. O cérebro é um computador altamente complexo, não-linear e paralelo. Ele tem a capacidade de organizar seus componentes estruturais, conhecidos como neurônios assim como realizar certas operações (reconhecimento de padrões, percepção e controle motor) muito mais rápido que o computador digital mais avançado existente nos dias de hoje. O cérebro humano possui algo diferencial dos computadores, a possibilidade de aprender e evoluir com base em sua experiência. Uma RNA é uma máquina com o objetivo de simular o modo em que o cérebro realiza uma tarefa em particular ou uma função de interesse.

Os neurônios artificiais, que compõem a unidade fundamental de processamento de informação de uma RNA possuem três elementos básicos:

- Conexões de entrada, que são ponderadas por um peso sináptico e posteriormente conectadas ao neurônio,
- Combinador linear, responsável pela soma da multiplicação dos valores de entrada Xi pelos seus respectivos pesos Wi e do limiar de ativação θ, resultando no potencial de ativação u,

$$u = \sum_{i=1}^{N} Wi \times Xi - \theta \tag{2.9}$$

- De acordo com [2], função de ativação, que avalia o potencial de ativação gerado pelo combinador linear e calcula o sinal de saída do neurônio. As funções mais utilizadas são a degrau, descrita pela Equação 2.10

$$g(u) = \begin{cases} 1, \text{ se } u \geq 0 \\ 0, \text{ se } u \leq 0 \end{cases} \tag{2.10}$$

  e uma sigmoidal, como a da Equação 2.11.

$$g(u) = 1/(1 + e^{-u}) \tag{2.11}$$

O tipo de RNA estudada neste trabalho é a *Perceptron* Multicamadas (*Multi-Layer Perceptron* - MLP), exemplificada pela Figura 2.2, que consiste tipicamente em um conjunto de unidades sensoriais que constituem a camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Ele é uma rede cujo aprendizado é geralmente feito através de um algoritmo de retropropragação do erro.

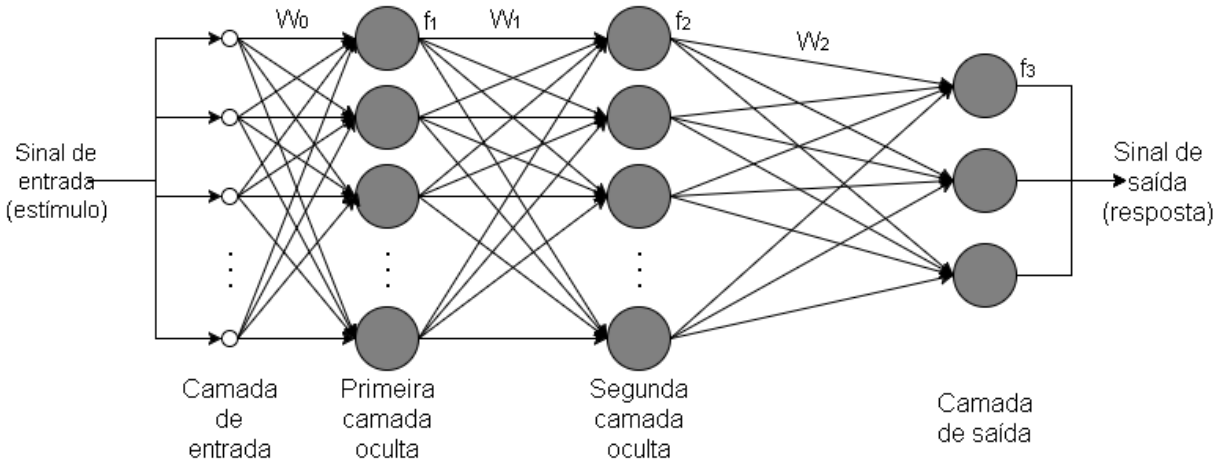

Figura 2.2: Exemplo de uma rede perceptron multicamadas, adaptado de [2].

A retropropagação de erro é um tipo de treinamento exaustivo que pode demorar bastante tempo dependendo do critério de parada e que consiste em:

1. Inicialização dos pesos da rede de forma aleatória ou seguindo algum método,
2. Processamento direto, as entradas passam pela rede contendo os novos pesos e é calculado o erro,
3. Teste de parada, o critério de parada adotado é testado, caso não tenha sido preenchido, termine.
4. Retropropagação, volta-se ao passo 2.

A partir dos pesos definidos, pode-se utilizar a RNA configurada para estimação de novos valores. As redes neurais que utilizam esse algoritmo de retropropagação de erro podem ser vistas como incógnitas, pois o resultado varia conforme o método utilizado e conforme a execução mas também porque pode-se não saber exatamente como a rede chega a um determinado resultado.

Basicamente, a aprendizagem por retropropagação de erro consiste de dois passos através das diferentes camadas de rede: um passo para frente, a propagação, e um passo para trás, a retropropagação. No passo para frente, um padrão de atividade (vetor de entrada) é aplicado aos nós sensoriais da rede e seu efeito se propaga através da rede. Durante o passo de propagação, os pesos sinápticos da rede são todos fixos. Durante o passo para trás, por outro lado, os pesos sinápticos são todos ajustados de acordo com uma regra de correção de erro. Especificamente, a resposta real da rede é subtraída de uma resposta desejada (alvo) para produzir um sinal de erro. Este sinal de erro é então propagado para trás através da rede, contra a direção das conexões sinápticas - vindo daí

o nome de "retropropagação de erro". Os pesos sinápticos são ajustados para fazer com que a resposta real da rede se mova para mais perto da resposta desejada, em um sentido estatístico. O algoritmo de retropropagação de erro é também referido na literatura como algoritmo de retropropagação. O processo de aprendizagem realizado com o algoritmo é chamado de aprendizagem por retropropagação [2].

O algoritmo de retropropagação utilizado neste trabalho foi o de Levenberg-Marquardt, uma técnica usada para resolver problemas de métodos dos mínimos quadrados não lineares [47].

Capítulo 3

Metodologia

O método apresentado para reconhecimento automático de traços de personalidade a partir de vídeos está representado pela Figura 3.1.

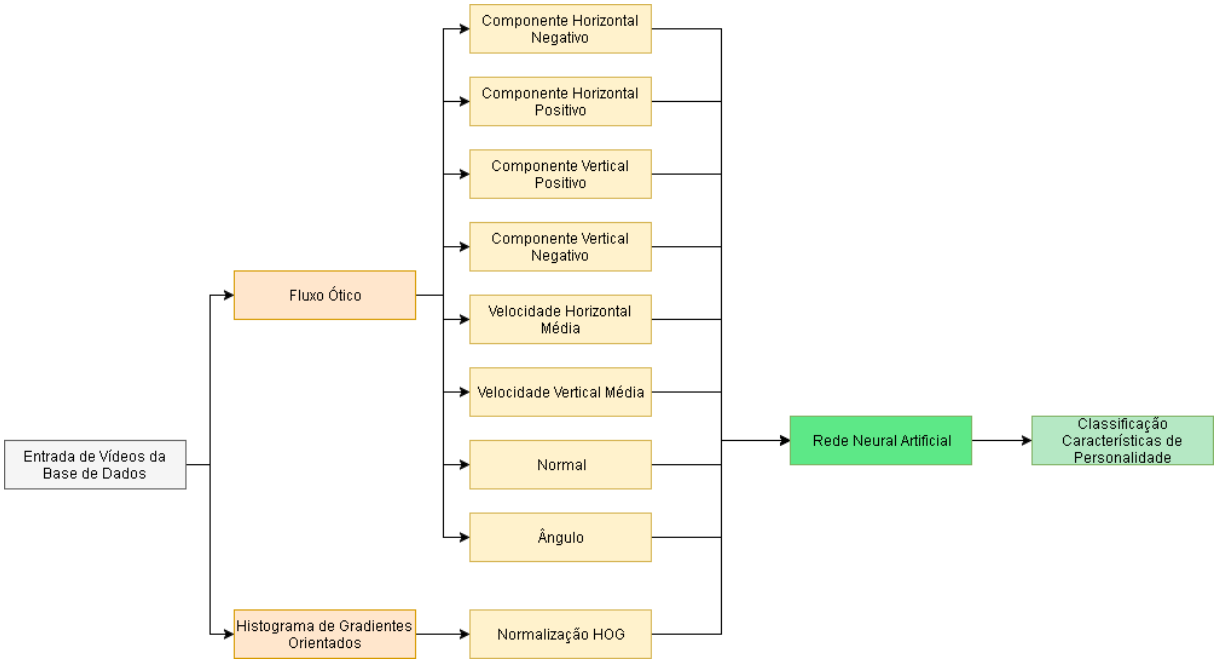

Figura 3.1: Fluxograma da Metodologia Proposta.

3.1 Base de dados

A base de dados utilizada foi a disponibilizada pelo desafio do *"Chalearn Looking at People Challenge and workshop @ ECCV 2016"*, que consiste em 10.000 recortes de aproximadamente 15 segundos de vídeos extraídos de mais de 3.000 vídeos de pessoas de gênero, idade, nacionalidade e etnicidade diferentes, falando em inglês, nativo ou não, frente à

uma câmera. Os vídeos possuem resolução e taxa de quadros variáveis, foram retirados da página de vídeos públicos do YouTube e avaliados pela *Amazon Mechanical Turk* (AMT) em relação aos traços de personalidade, definidos como *Big Five*. Essas avaliações correspondem a números decimais entre 0 e 1 para cada um dos traços e serão os rótulos utilizadas para a classificação.

Algumas restrições foram feitas para os vídeos:

- Apenas uma pessoa presente no plano de frente,
- Imagem e áudio de boa qualidade,
- Vídeos com apenas a língua inglesa sendo falada,
- Pessoas com idade acima de 13 anos,
- Pouco movimento da câmera (permitido mudança de fundo, porém evitar borrões),
- Sem conteúdo adulto ou violento (porém pode-se falar sobre o assunto),
- Sem nudez,
- Pode haver pessoas no fundo da imagem (sem barulho e baixa resolução das faces),
- Sem propagandas,
- Evitar cortes de áudio ou vídeo (mudanças bruscas).

Os vídeos que melhor se adequaram às características desejadas foram vídeos de perguntas e respostas, pois, no geral, são vídeos com poucas pessoas presentes, conversa abundante com a câmera, pouco movimento no fundo e voz clara.

Dos 10.000 vídeos disponibilizados pela base de dados, utilizou-se os 6.000 cujos rótulos foram divulgadas com os valores do *Big Five*, métrica de avaliação de personalidade que será descrita na próxima subseção. Algumas amostras da base de dados encontram-se na Figura 3.2.

3.1.1 Big five

O *Big Five*, também conhecido como o Modelo de Cinco Fatores, é um termo da psicologia que se refere a análise de personalidade de um indivíduo a partir de questionários [48] que determinam cinco principais características de um indivíduo, a saber:

- Neuroticismo (*neuroticism*), ou instabilidade emocional, representa a tendência para emoções negativas. Um nível alto de neuroticismo tende a representar pessoas muito emotivas, vulneráveis ao estresse. Níveis baixos de neuroticismo tendem a representar indivíduos calmos e emocionalmente estáveis.

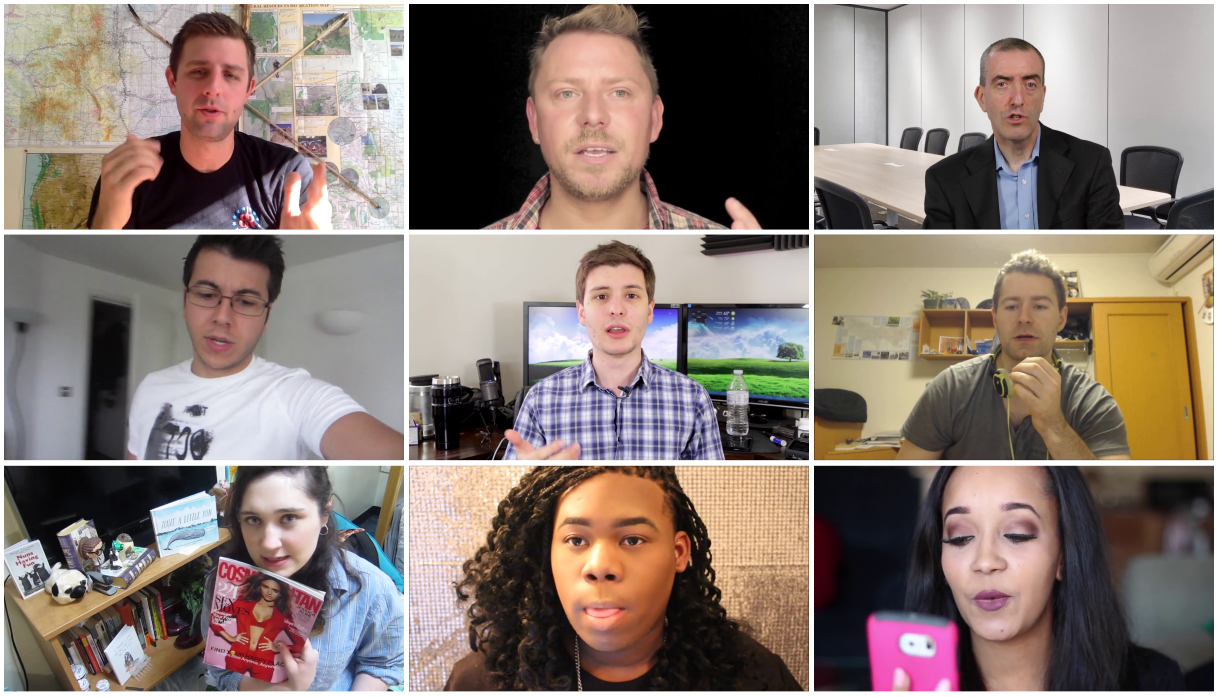

Figura 3.2: Amostras da base de dados.

- Extroversão (*extraversion*) representa tendência a procurar por estimulação e companhia dos outros, procurar contato com o mundo exterior. Já introvertidos preferem sua privacidade e se sentem bem consigo mesmos, sem necessidade de alguém exterior.
- Amabilidade (*agreeableness*) representa a tendência a se preocupar com o bem-estar dos outros, já pessoas que apresentam níveis baixos de amabilidade possuem como característica botar o interesse próprio acima de uma boa relação com os outros.
- Escrupulosidade (*conscientiousness*) representa um comportamento planejado no lugar de espontâneo. Indivíduos com níveis altos nessa característica tendem a mostrar autodisciplina.
- Abertura para a experiência (*openness to experience*) distingue pessoas imaginativas das pessoas convencionais. Altos níveis desse traço geralmente estão presentes em pessoas intelectualmente curiosas e apreciadores de arte. Já níveis baixos caracterizam pessoas preferem o simples, claro e óbvio ao ambíguo e complexo.

Essas características também são organizadas e definidas pela Figura 3.3. Sendo avaliadas, elas possuem utilizações bastante relevantes como avaliar candidatos para uma vaga de emprego, conseguindo determinar qual seria o candidato que tende a ser mais apto, adequado para a função requerida.

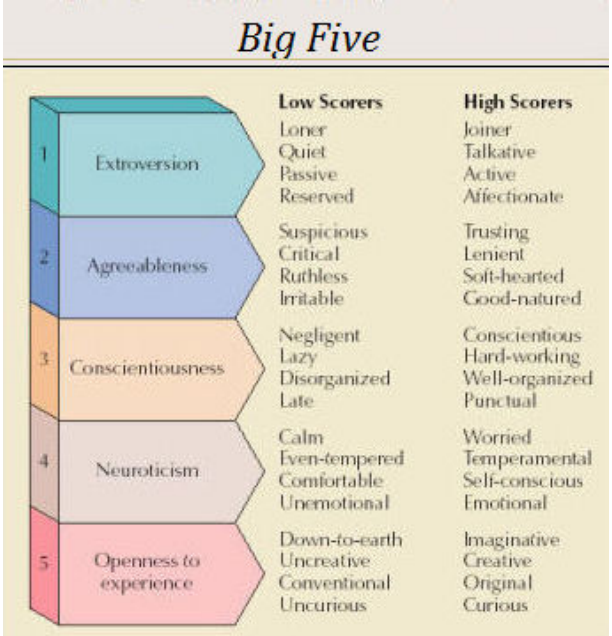

Figura 3.3: Resumo do *Big Five*, retirado de [3].

3.1.2 Fluxo Ótico

O Fluxo Ótico aplicado foi o de Horn e Schunck [23], descrito na Seção 2.2. A partir desta metodologia utilizada para a obtenção do fluxo ótico, extraiu-se as componentes horizontais e verticais, a velocidade média horizontal e vertical, o valor médio da normal e o ângulo médio entre a componente horizontal e vertical para cada pixel da imagem, respectivamente definidos pelas variáveis $Velx$, $Vely$, $Medx$, $Medy$, $Norm$ e Ang, para cada vídeo.

Para cada vídeo, as componentes $Velx$ e $Vely$ foram separadas em valores positivos e negativos e foi feito um somatório do número de ocorrências para cada caso. Ou seja, se $Velx$ é positivo, soma-se 1 à variável contadora de componentes de velocidade vertical positivo, $VelxP$, caso contrário, soma-se 1 à variável contadora de componentes de velocidade vertical negativo, $VelxN$. O mesmo é feito para as componentes horizontais, $VelyP$ e $VelyN$. Feito isso para um vídeo inteiro, normalizou-se esses valores, resultando em quatro características que serão utilizadas como base para a etapa de classificação, C_1, C_2, C_3 e C_4, de acordo com as Equações 3.1 a 3.4, com valores decimais entre 0 e 1.

$$C_1 = \frac{VelxP}{VelxP + VelxN + VelyP + VelyN} \tag{3.1}$$

$$C_2 = \frac{VelxN}{VelxP + VelxN + VelyP + VelyN} \tag{3.2}$$

$$C_3 = \frac{VelyP}{VelxP + VelxN + VelyP + VelyN} \tag{3.3}$$

$$C_4 = \frac{VelyN}{VelxP + VelxN + VelyP + VelyN} \tag{3.4}$$

Para a velocidade média, somou-se o valor de $Velx$ e $Vely$, separadamente, para cada pixel e para cada quadro, em seguida esse valor foi dividido pelo número de pixels, sendo a *altura* pela *largura* e pelo número de quadros, $numDeQuadros$, em cada vídeo, resultando em duas características para a etapa de classificação, C_5 e C_6, de acordo com as Equações 3.5 e 3.6.

$$C_5 = \frac{\sum Velx}{altura * largura * numDeQuadros} \tag{3.5}$$

$$C_6 = \frac{\sum Vely}{altura * largura * numDeQuadros} \tag{3.6}$$

Para o valor médio da normal, adicionou-se o quadrado de $Velx$ ao quadrado de $Vely$, para cada pixel e para cada quadro do vídeo, para em seguida extrair a raiz quadrada e dividir pelo número de pixels e pelo número de quadros, resultando na característica C_7 para a etapa de classificação e exemplificado pela Equação 3.7.

$$C_7 = \frac{\sum\sqrt{Velx^2 + Vely^2}}{altura * largura * numDeQuadros} \tag{3.7}$$

Como última característica extraída do fluxo ótico, C_8 foi calculado o ângulo médio da normal, a partir da soma da arco tangente de $Vely$ sobre $Velx$, para cada pixel e quadro, divido pelo número de pixels e número de quadros, dado na Equação 3.8.

$$C_8 = \frac{\sum(\arctan Vely/Velx)}{altura * largura * numDeQuadros} \tag{3.8}$$

3.1.3 Histograma de Gradientes Orientados

A partir do método de extração de histograma de gradientes orientados, descrito na Seção 2.3, para cada quadro da imagem, obteve-se um vetor de 1 por N, onde N é o tamanho da característica do HOG e o valor do vetor corresponde a diferença dos *bins*. Os valores desse vetor foram somados, juntamente com os valores encontrados para os outros quadros do vídeo, resultando em um valor único. Para esse valor, foram propostos quatro métodos para sua normalização.

Como a resolução e a taxa de quadros da base de dados é variável de vídeo a vídeo, essas informações foram as utilizadas para realizar a normalização. Para sua melhor definição,

foram propostos 4 modos diferentes para normalizar, representados pelas Equações 3.9 a 3.12. Os dados obtidos encontram-se na Figura 3.4.

$$C_9 = \frac{\sum_1^{ndequadros} \sum_1^N HOG}{(altura * largura)} \tag{3.9}$$

$$C_9 = \frac{\sum_1^{ndequadros} \sum_1^N HOG}{(altura + largura)} \tag{3.10}$$

$$C_9 = \frac{\sum_1^{ndequadros} \sum_1^N HOG}{taxaDeQuadros * (altura * largura)} \tag{3.11}$$

$$C_9 = \frac{\sum_1^{ndequadros} \sum_1^N HOG}{taxaDeQuadros * (altura + largura)} \tag{3.12}$$

Vê-se os resultados são bem próximos, porém o melhor, em média, é quando divide-se o valor encontrado na soma do HOG pela taxa de quadros multiplicada pela soma da altura e largura da imagem. Esse resultado é representado pela Equação 3.12 e a partir dele é obtida a 5a característica C_9 para a etapa de classificação.

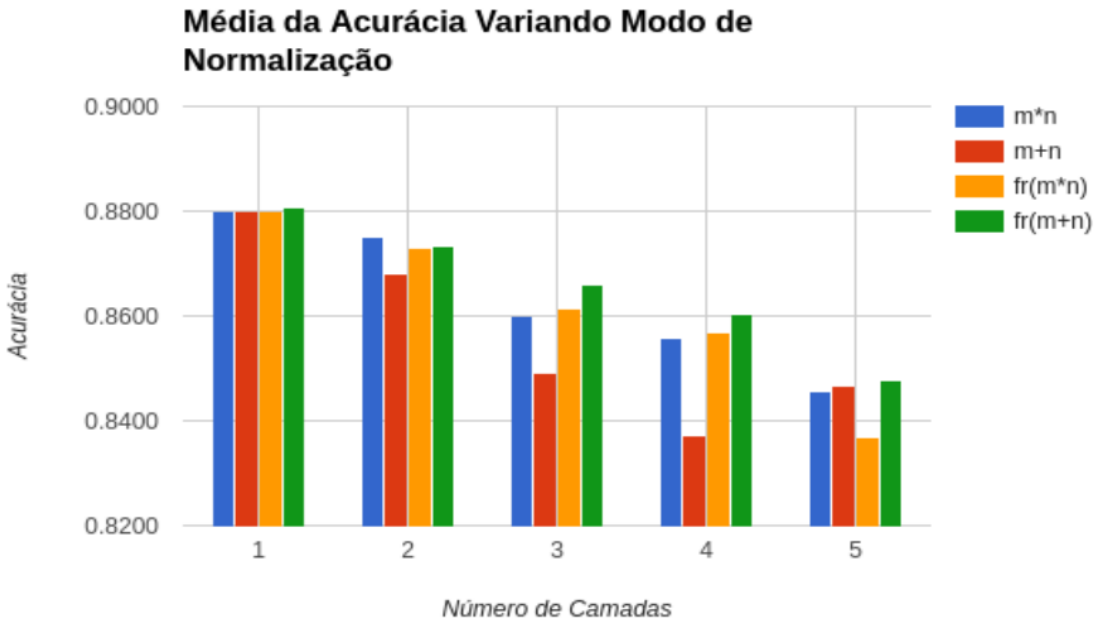

Figura 3.4: Gráfico da média da acurácia variando número de camadas e modo de normalização do HOG.

3.1.4 Rede Neural Artificial

As nove características C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8 e C_9 encontradas nas etapas anteriores foram empregadas para treinamento da rede neural *perceptron* multicamadas descrita na Seção 2.4, Capítulo 2, usando a função de Levenberg-Marquardt [49] como função de ativação.

Para implementação da rede neural, foi necessário definir o número de camadas ocultas e neurônios mais adequado para o problema. Para essa definição, foram utilizados 1000 vídeos da base de dados, separados, conforme a literatura [2] indica, em 700 para treinamento e 300 para validação. O método de treinamento usado foi o de teste manual, testando-se a acurácia para cada caso testado

Definição do Número de Camadas Ocultas

Primeiramente, para determinar o número de camadas ocultas, foram utilizados variou-se o número de camadas de 1 a 5 com 10 neurônios cada e calculou-se a acurácia em cada caso para definir um número adequado. Fez-se utilizando diversas normalizações para o valor do HOG. Os resultados alcançados encontram-se nos gráficos das Figuras 3.5, 3.6, 3.7 e 3.8, onde fr corresponde a taxa de quadros, m corresponde a altura e n a largura do vídeo.

Desses gráficos, percebe-se uma leve decaída, em média de 0.024 da média geral, para cada aumento de camadas, na acurácia dos resultados conforme aumenta-se o número de camadas e que os melhores resultados foram obtidos utilizando uma camada oculta. Essa diferença, mesmo que bem pequena, foi a base para definição do número de camadas, estabelecido como 1, mais adequado para o problema.

Definição do Número de Neurônios

Tendo o número de camadas definido, variou-se o número de neurônios para definir um que gere melhor acurácia para o problema, conforme o gráfico da Figura 3.9.

Variou-se de 1 a 73 neurônios e nota-se uma decaída na acurácia conforme o número aumenta, com o pico de acurácia localizado entre 1 e 10 neurônios. Parou-se em 73 neurônios pois notou-se uma queda na acurácia conforme aumentava-se o número de neurônios. Para definir o melhor número dentro dessa variação realizou-se o treinamento e validação da rede neural 10 vezes para cada número de neurônio entre 1 e 10, utilizando uma camada oculta para obter uma acurácia média e descobrir com quantos neurônios obtém-se o melhor resultado no geral, conforme mostrado no gráfico da Figura 3.10.

A partir do gráfico, fixa-se o número de neurônios em 2, devido à acurácia média obtida nestes testes ser maior do que a das outras configurações. Entretanto, este teste

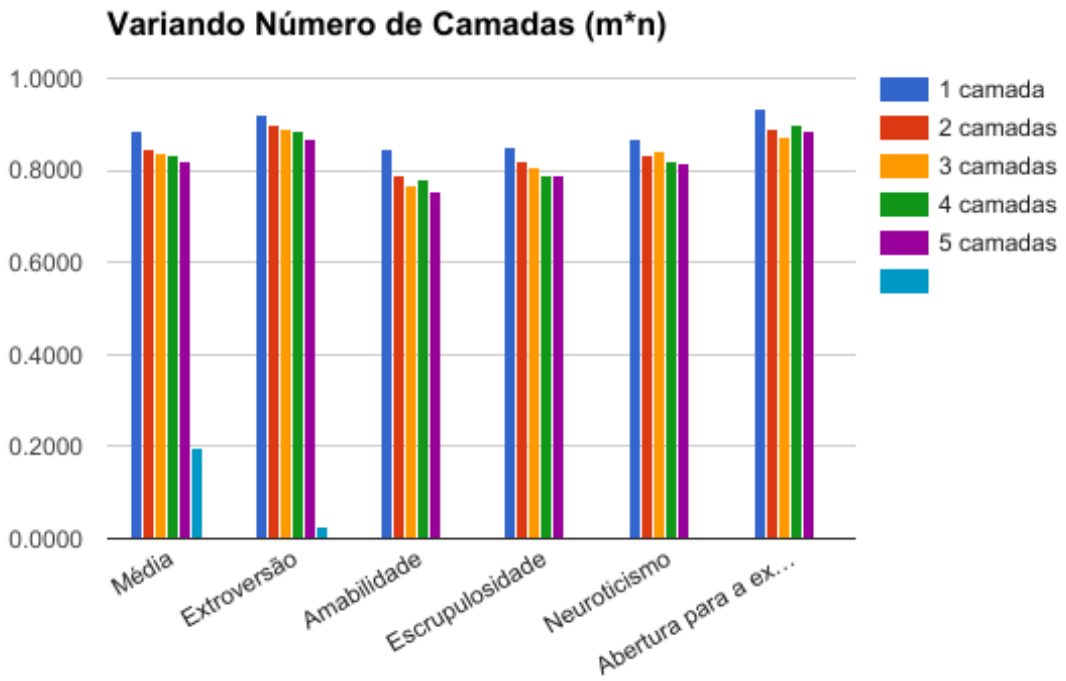

Figura 3.5: Gráfico variando camadas normalizando HOG dividindo-o pela resolução (m*n).

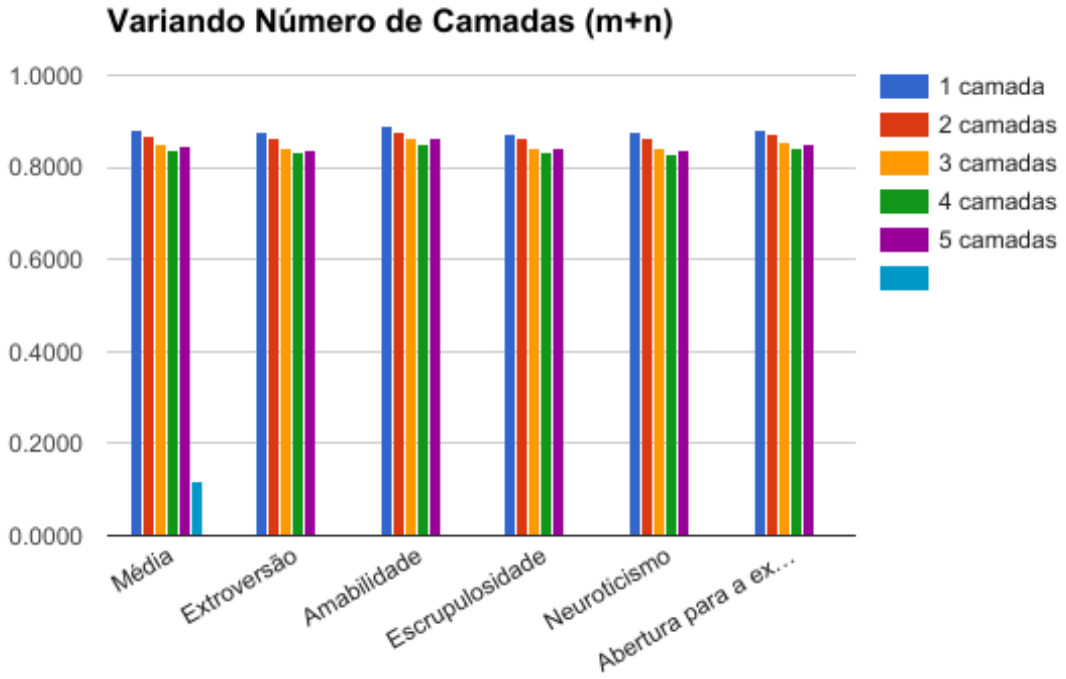

Figura 3.6: Gráfico variando camadas normalizando HOG dividindo-o por m+n.

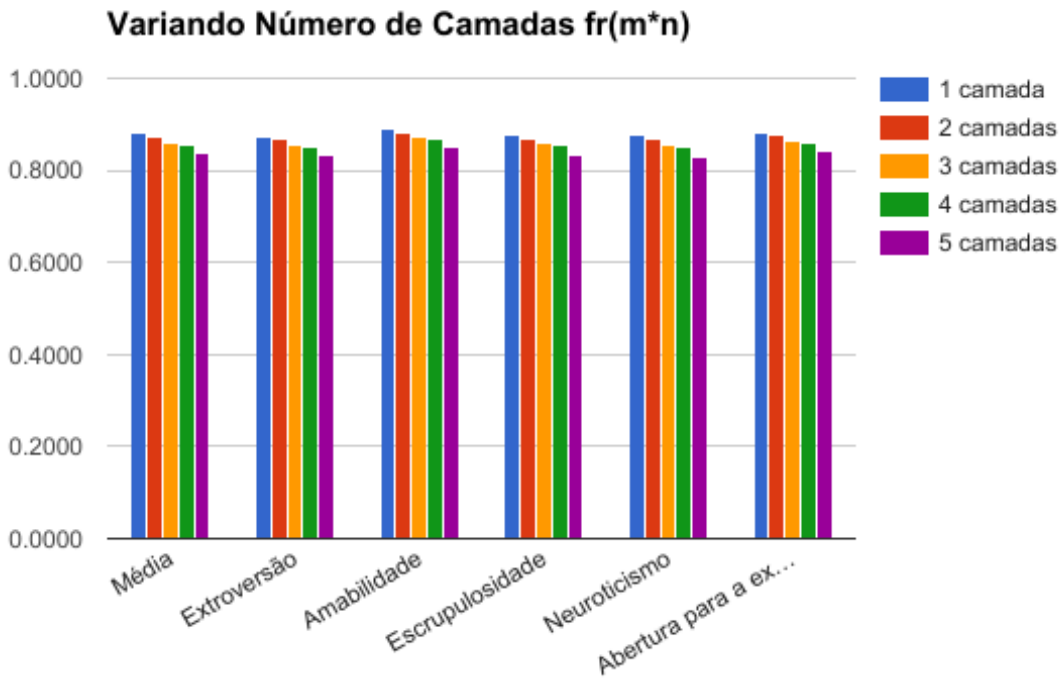

Figura 3.7: Gráfico variando camadas normalizando HOG dividindo-o por fr(m*n).

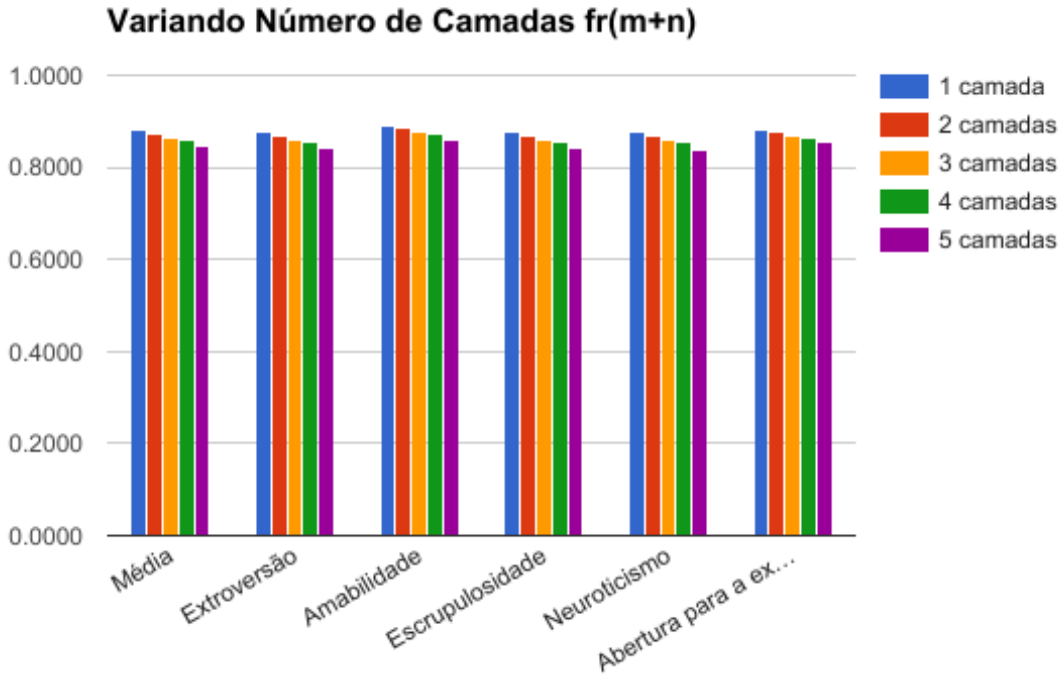

Figura 3.8: Gráfico variando camadas normalizando HOG dividindo-o por fr(m+n).

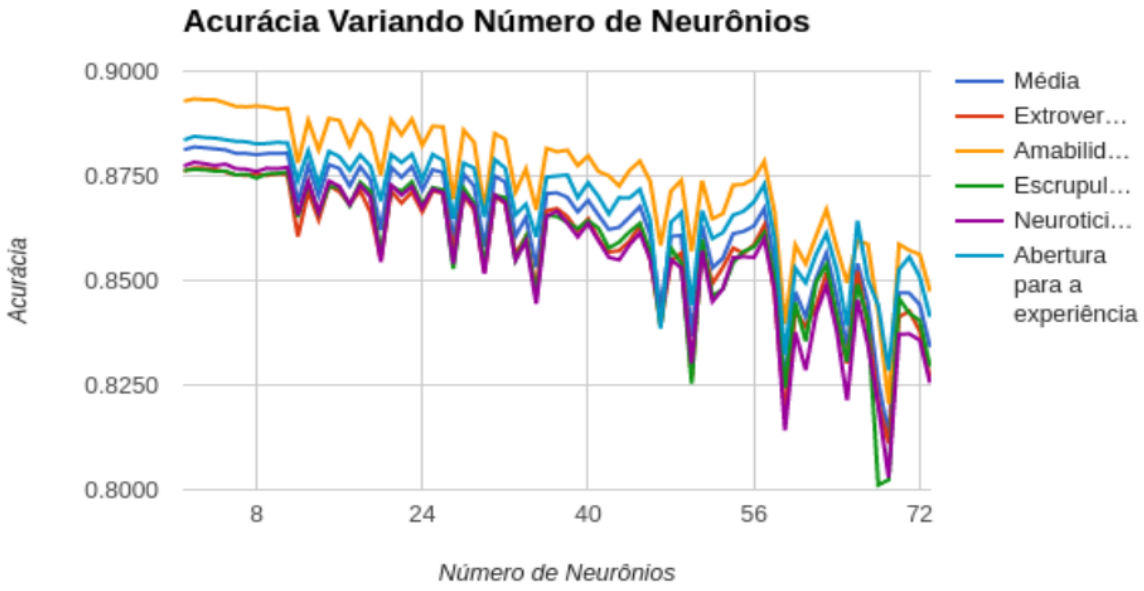

Figura 3.9: Gráfico variando número de neurônios.

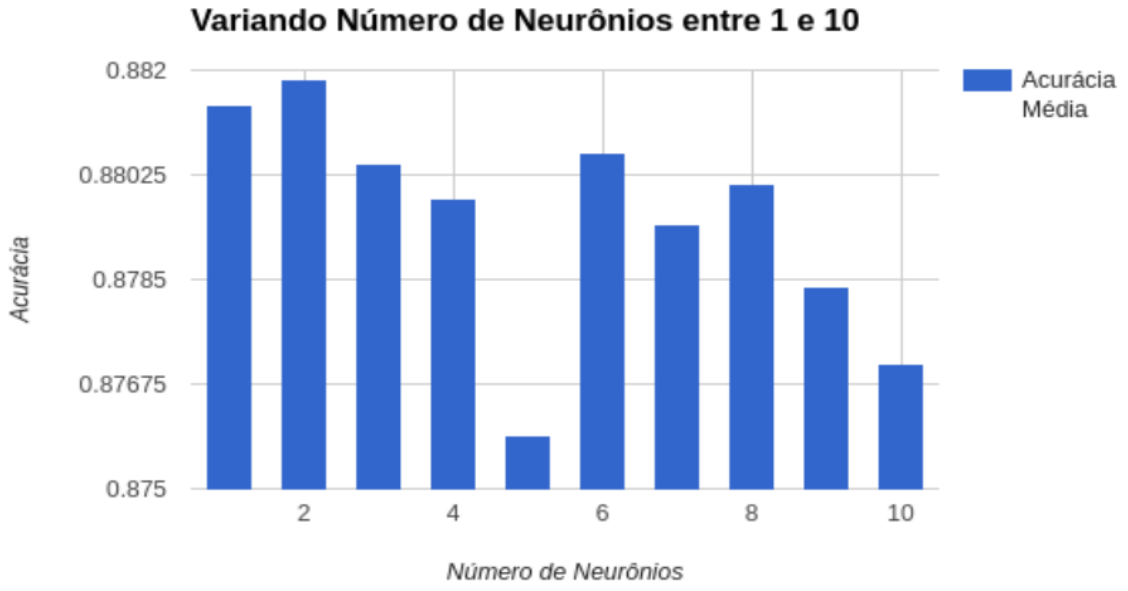

Figura 3.10: Gráfico da média da acurácia variando número de neurônios entre 1 e 10.

não garante que será a melhor forma de se obter os resultados, sendo assim reavaliado no capítulo de resultados.

3.1.5 Métrica de Avaliação

A métrica utilizada para computação dos resultados foi a acurácia, encontrada de acordo com a Equação 3.13, onde *Label* corresponde ao resultado esperado, *Saída* corresponde ao resultado encontrado e $NúmeroDeVídeos$ corresponde ao número de vídeos utilizados na amostra.

$$Acc = 1 - \frac{\sum |Label - Saída|}{NúmeroDeVídeos} \tag{3.13}$$

A equação corresponde a 1 menos a distância média entre o valor encontrado pelo trabalho e o valor esperado. Os valores de saída são decimais entre 0 e 1, portanto, não tem como ser preciso e sim acurado. Tenta-se chegar o mais perto possível do valor esperado, ou seja, obter-se uma acurácia de valor ideal igual a 1, o que significaria que o resultado encontrado é exatamente igual ao resultado esperado. Portanto, quanto mais perto de 1 for o valor encontrado, mais acurado será o resultado.

Capítulo 4

Resultados

O código de Fluxo Ótico foi desenvolvido usando a biblioteca *Open Computer Vision*[50] (OpenCV) para processamento de imagens e o código de HOG e rede neural foram desenvolvidos a partir da plataforma do Matlab.

Como critério de parada para o treinamento da rede neural, foi utilizado um valor máximo de 1000 épocas, que foi o critério que definiu a parada na grande parte dos casos, um valor mínimo de 0.00 para o desempenho, um valor mínimo de 1.00×10^{-07} para o gradiente, que foi o outro critério que definiu a parada, porém em número bem menor de casos, e um valor máximo de 1.00×10^{10} para o μ, que é uma variável que é aumentada de um valor fixo, *mu_inc*, até que a mudança resulte em uma melhora na performance, obtido esse resultado, a variável é decrescida de outro valor fixo, *mu_dec* [51].

A arquitetura computacional usada para a execução do código possui as seguintes especificações:

- Memória RAM: 3.8 GB
- Processador: Intel Core i3-2328M CPU @ 2.20GHz x 4
- Sistema operacional: Linux Mint 17.3 Rosa 64-bit

4.1 Análise dos Resultados

A partir da técnica utilizada, foram testadas diversas configurações, como visto no gráfico da Figura 4.1, a fim de se determinar a divisão entre vídeos utilizados para treinamento e vídeos para validação. Percebe-se uma leve decaída da acurácia conforme diminui-se a quantidade de vídeos para treinamento e aumenta-se a quantidade para classificação, com um pico de 0.8824 na divisão de 70% dos 6000 vídeos para treinamento e 30% para validação. Dessa forma, essa foi a divisão definida.

Figura 4.1: Gráfico da média da acurácia variando porcentagem utilizada para treinamento e para validação.

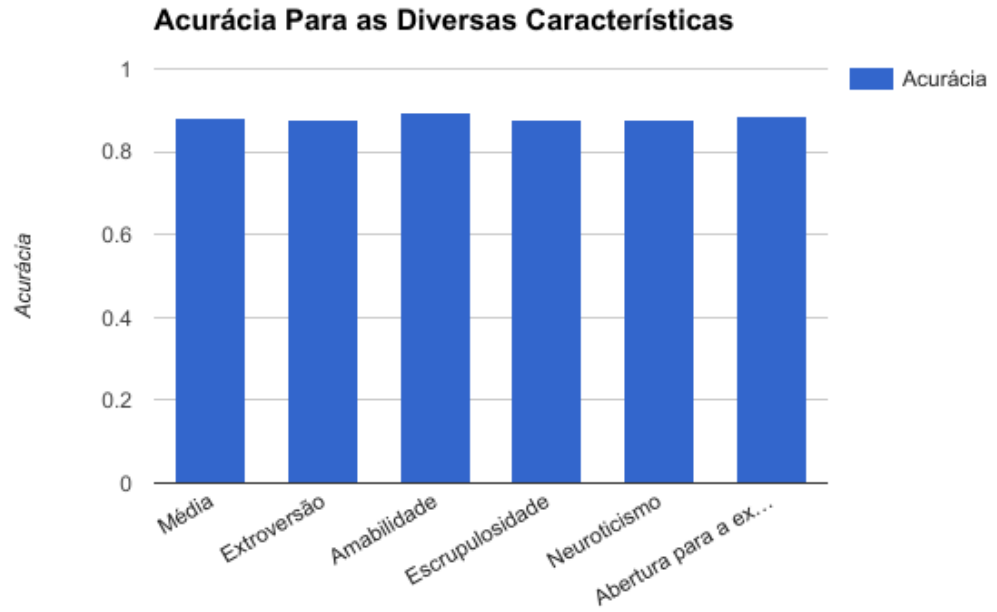

Figura 4.2: Gráfico do resultado de acurácia final.

Inicialmente, as características C_5, C_6, C_7 e C_8 não haviam sido extraídas e acreditava-se que os resultados encontrados eram relevantes, pois a acurácia estava alta. Porém, após

uma observação mais apurada nos resultados obtidos, percebeu-se que a rede neural não respondia corretamente e os valores não estavam variando, como pode-se ver na Figura 4.3

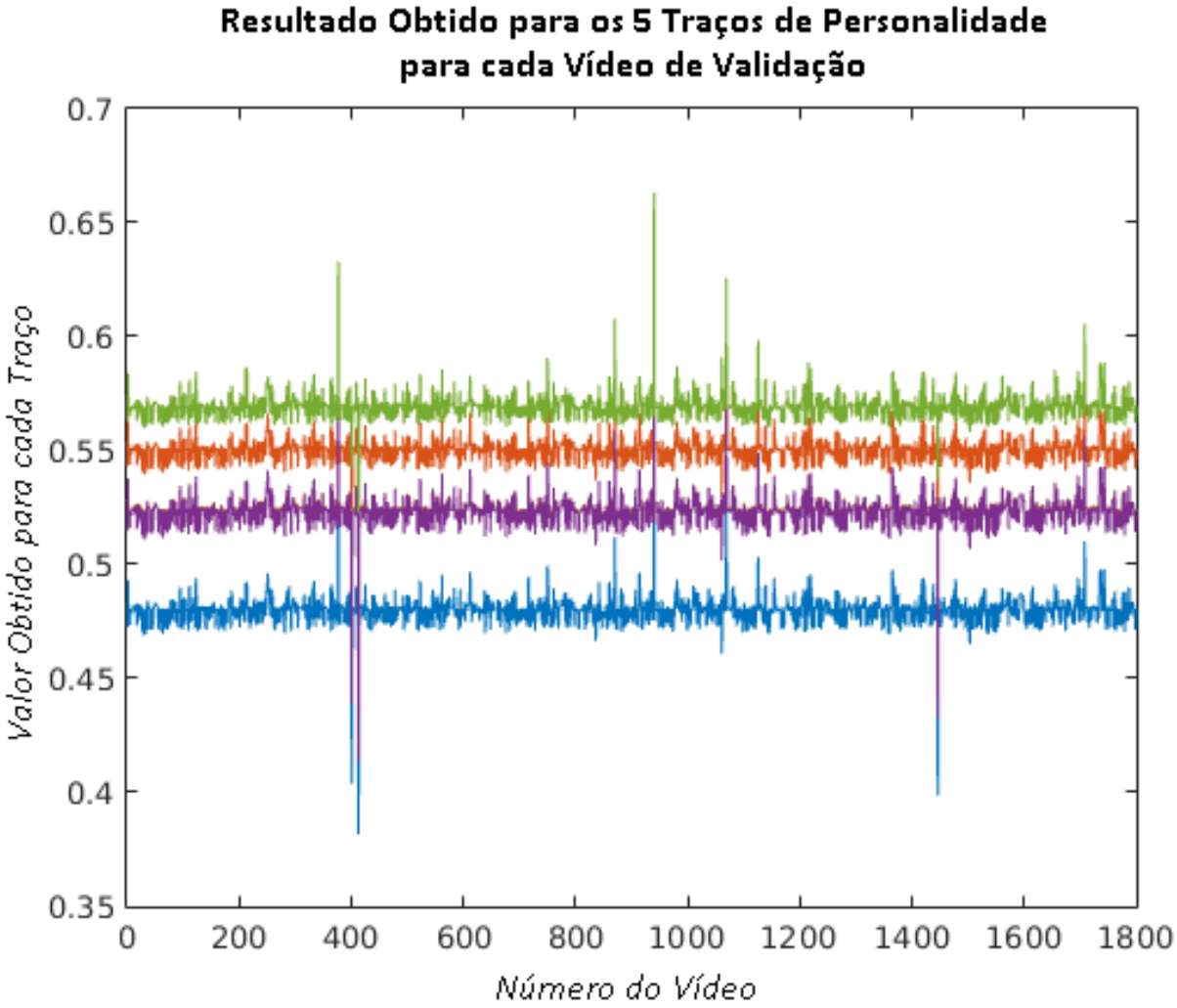

Figura 4.3: Gráfico dos resultados da configuração inicial. Cada cor corresponde a um traço de personalidade.

Após essa observação, foi necessário obter novas características e procurar meios distintos para testar a rede. Os cenários incluídos com as novas características serão apresentados nas Subseções 4.1.1 e 4.1.2.

4.1.1 Cenário 1

Neste cenário, será isolada apenas uma das 5 características, considerando essa como a dominante para avaliar a capacidade de treinamento mediante ao tipo de saída esperada.

Para cada parte do cenário, será utilizada a configuração de camadas e neurônios determinada na Subseção 3.1.4. Como neste caso, os resultados agora não são mais decimais, a acurácia foi calculada de acordo com a Equação 4.1

$$Acc = \frac{TP + TN}{P + N} \tag{4.1}$$

onde TP corresponde aos verdadeiros positivos, TN corresponde aos verdadeiros negativos, P corresponde aos positivos e N aos negativos.

Parte 1

Em um primeiro momento, as saídas foram modificadas para que a rede neural definisse o traço de personalidade principal presente no vídeo, isto é, o maior valor encontrado nos 5 traços foi definido como +1 e o resto dos valores foram modificados para -1. Por exemplo, em um vetor cujos valores dos traços de características de saída esperado fossem [0.35 0.48 0.21 0.87 0.63], o vetor resultante de saída da rede neural utilizada seria [-1 -1 -1 +1 -1].

A partir dessa nova saída, treinou-se a rede neural novamente e a rede aparentava mostrar uma classificação, com uma precisão de 0.3608, porém, percebeu-se que a rede apontava 77,11% das vezes para o mesmo traço e dois traços não foram classificados, não validando a técnica, como visto na Figura 4.4.

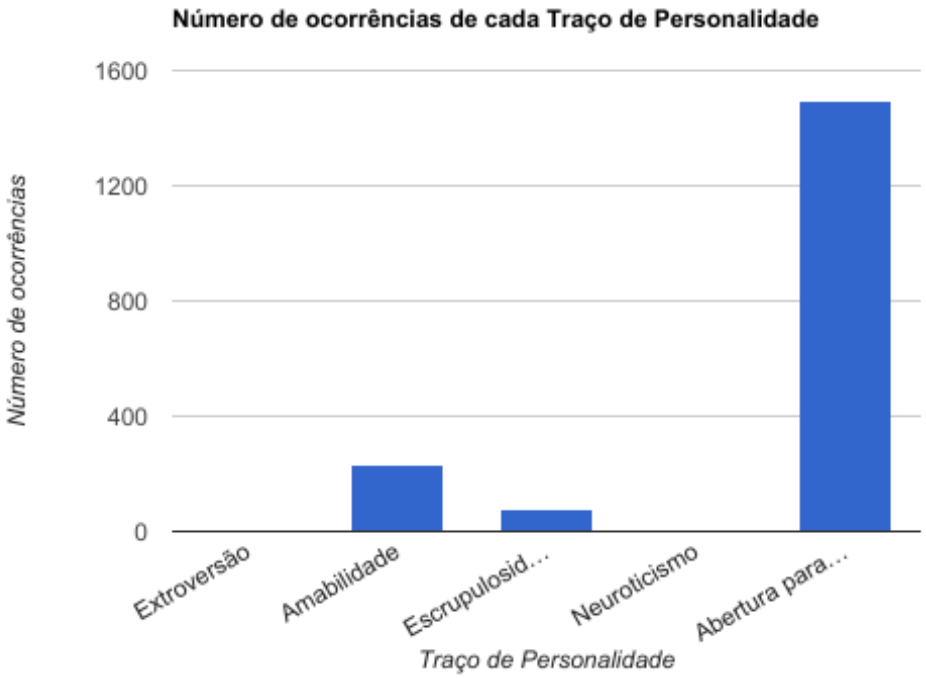

Figura 4.4: Gráfico dos resultados do cenário 1.

Parte 2

Em uma segunda tentativa, as saídas foram novamente modificadas, porém, ao invés de definir como +1 o traço de personalidade principal e -1 os outros, foi definido como 1 o principal e 0 o restante. Resultando em, para um vetor cujos valores dos traços de características fosse [0.35 0.48 0.21 0.87 0.63], um novo vetor que seria [0 0 0 1 0], como apresentado anteriormente na Parte 1.

Treinada novamente a rede, obteve-se o mesmo resultado anterior, apontando predominantemente para o mesmo traço de personalidade. Uma explicação é que a partir das características aqui levantadas, a rede neural não conseguiu alcançar variação, como verificado pela Figura 4.5.

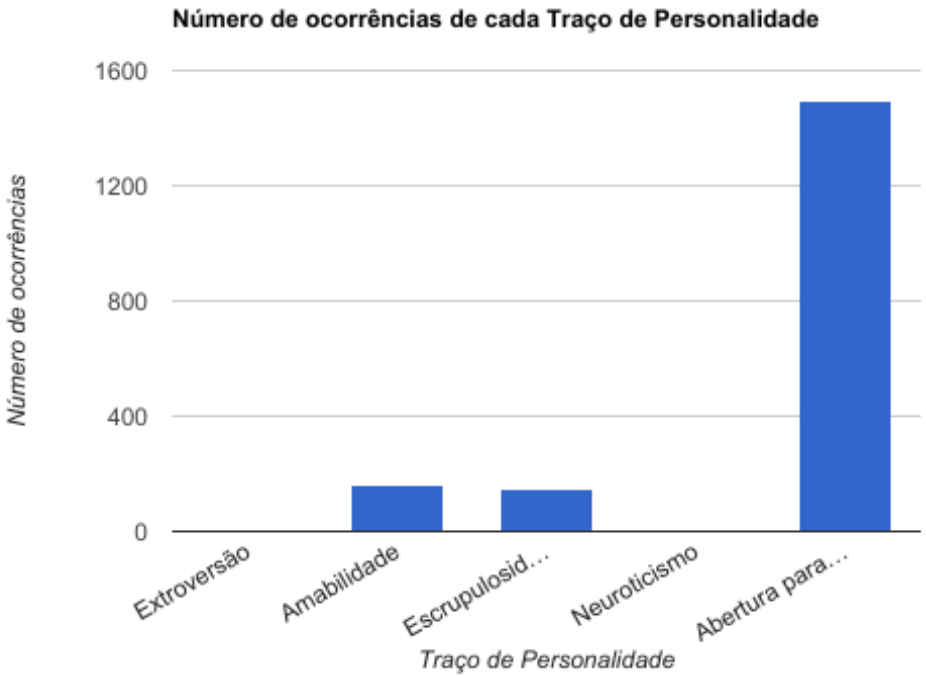

Figura 4.5: Gráfico dos resultados do cenário 1.

Parte 3

Tentou-se, ainda, a partir da modificação das partes 1 e 2, separar cada traço de personalidade individualmente para a confecção de 5 redes paralelas, referentes a cada um. Ou seja, uma rede neural era treinada para a amabilidade, uma segunda rede para a escrupulosidade, uma terceira para extraversão, uma quarta para neuroticismo e uma última para abertura para a experiência.

O resultado parecido como o restante dos casos, na combinação das saídas das 5 redes diferentes, observou-se que prevalece apenas um traço de personalidade dominante, porém neste caso, não houve nenhuma ocorrência de saída resultando em outro traço, como nota-se pela Figura 4.6.

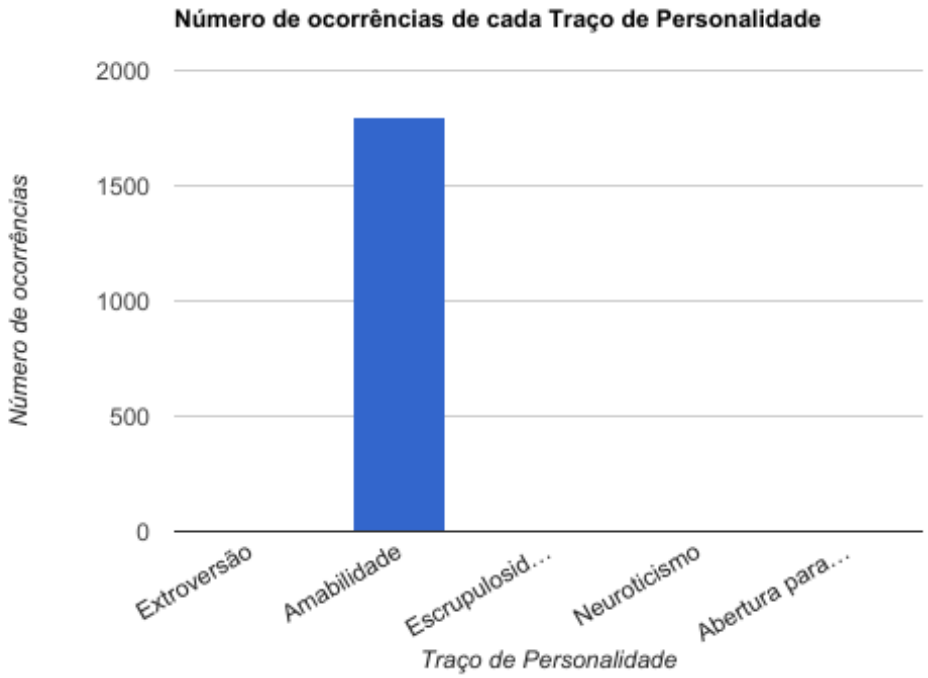

Figura 4.6: Gráfico dos resultados do cenário 1. Cada cor corresponde a um traço de personalidade.

4.1.2 Cenário 2

Uma explicação para a falta de sucesso Cenário 1 é que a determinação de um traço de personalidade dominante não estava possibilitando a classificação deste, a partir das características encontradas. Portanto, voltou-se para a abordagem original, de apenas uma rede e as saídas originais, porém agora com as novas características definidas.

Os resultados foram mais convincentes que os anteriores, e, mesmo a acurácia média encontrada sendo parecida com a encontrada previamente, sem as características adicionadas, as saídas começaram a variar muito mais, vide Figura 4.7, indicando uma resposta melhor da rede neural.

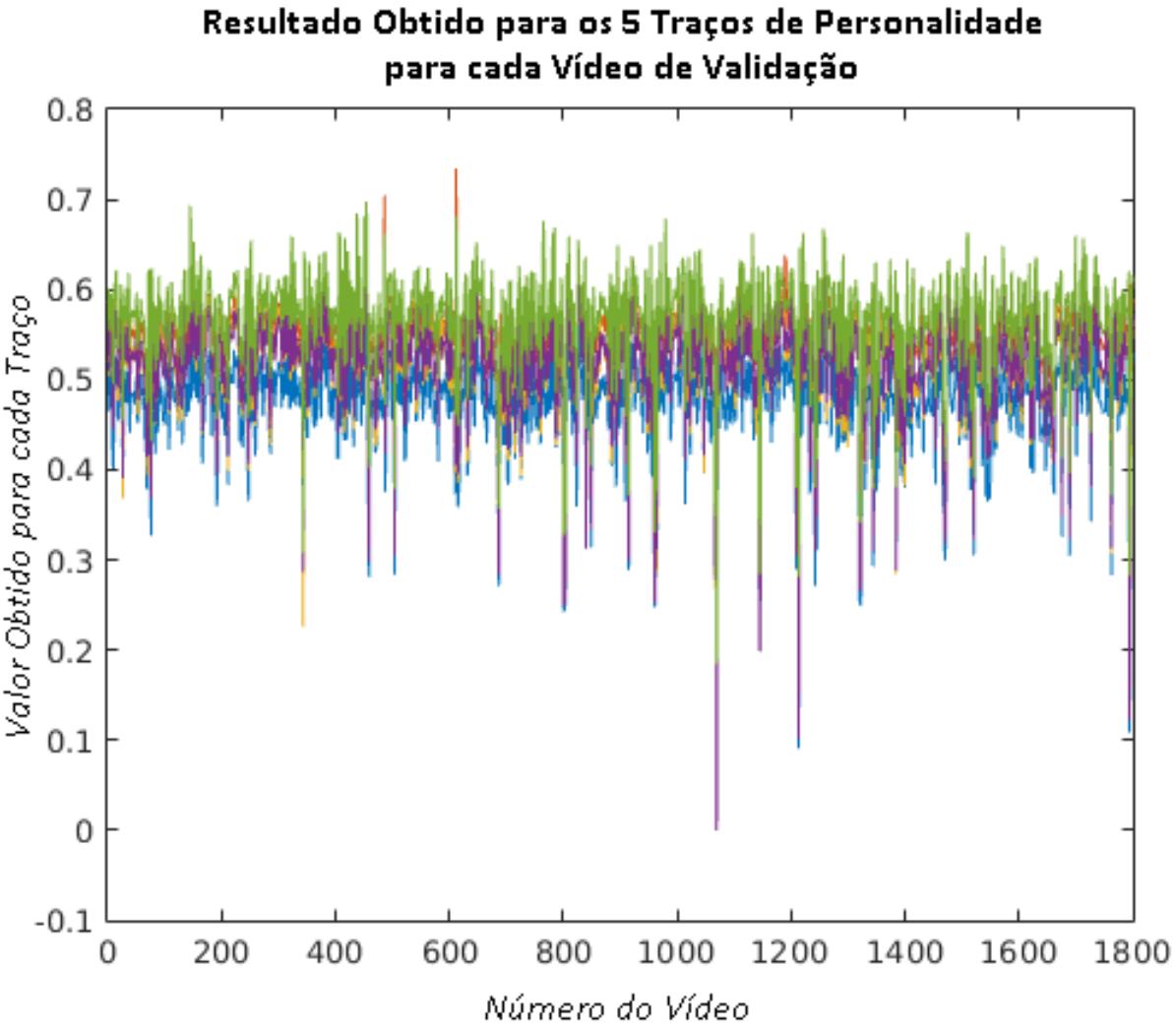

Figura 4.7: Gráfico dos resultados da configuração final.

4.1.3 Resultados Finais

A partir dos resultados obtidos nos cenários anteriores e das análises realizadas na Subseção 3.1.4, a configuração final da rede neural é a do Cenário 2 utilizando uma camada oculta de dois neurônios.

Para efeito de comparação com os resultados obtidos por equipes que participaram do desafio [3] de reconhecer automaticamente os traços de personalidade em vídeo, será utilizada a Equação 3.13 que descreve a acurácia, pois foi a equação utilizada para obtenção desses resultados no desafio.

Na Tabela 4.1, foram utilizados os resultados obtidos pelos dez primeiros colocados no desafio [4] na parte de aprendizagem, utilizando a acurácia, e foi adicionado o resultado obtido neste trabalho. Vale ressaltar que na maioria dos trabalhos, utilizou-se

informações de áudio para melhorar a acurácia e neste trabalho foram utilizadas somente informações de imagens. Os valores correspondem a acurácia para cada um dos traços de personalidade. Nota-se que os resultados obtidos neste trabalho são próximos aos melhores do desafio, constatando-se assim que os resultados deste trabalho são relevantes. Porém vale salientar que existe a possibilidade de a rede neural não estar respondendo, mesmo obtendo um resultado significativo, pois há chances de a rede estar expondo ruído como anteriormente, apenas variando melhor os valores de saída.

Vale ressaltar que os dados utilizados diferem, pois neste trabalho foi utilizada apenas a parte da base de dados de treinamento e, no desafio, esses resultados foram obtidos fundamentados da parte da base de dados de validação, cujos valores de saída não foram divulgados, portanto a tabela tem caráter apenas comparativo.

Tabela 4.1: Comparação do resultados deste trabalho com trabalho com os do desafio

	Usuário	Acurácia Média	Extroversão	Amabilidade	Escrupulosidade	Neuroticismo	Abertura
1	Arulkumar	0.912902 (1)	0.915924	0.916297	0.909889	0.910027	0.912374
2	vismay	0.912020 (2)	0.912856	0.91452	0.913203	0.907846	0.911675
3	frkngrpnr	0.911270 (3)	0.914486	0.912704	0.91289	0.906242	0.910026
4	flx	0.910226 (4)	0.911887	0.913819	0.90596	0.908365	0.911098
5	pandora	0.908182 (5)	0.907719	0.912539	0.903413	0.907272	0.909965
6	ashish	0.907549 (6)	0.908504	0.913363	0.903077	0.904102	0.908699
7	umuguc	0.906461 (7)	0.905721	0.911128	0.902121	0.907886	0.905451
8	stmater	0.892134 (8)	0.892285	0.902665	0.884485	0.887215	0.894021
9	tzzcl	0.887523 (9)	0.908509	0.880302	0.877144	0.891178	0.880482
10	Hao_Zhang	0.887424 (10)	0.885786	0.898298	0.876983	0.885531	0.890523
11	Este Trabalho	0.882143 (11)	0.877717	0.893539	0.877966	0.877303	0.883937

Capítulo 5

Conclusão e Tabalhos Futuros

Neste trabalho, a proposta inicial era formular uma técnica para reconhecimento automático de traços personalidade em vídeos, utilizando como referência os fatores de personalidade conhecidos na psicologia como *Big Five*. Para a realização deste reconhecimento, foram usadas cinco abordagens distintas que culminaram em uma metodologia final que é definida pela captura de características do fluxo ótico, assim como do histograma de gradientes orientados, para em seguida utilizá-las em um classificador que consiste em uma rede neural perceptron multicamadas.

Seguindo a metodologia proposta, alcançou-se uma acurácia de 0.8821 (88.21%) que, comparado com os resultados obtidos pelas outra técnicas no desafio [4], é um valor competitivo e mostra que a metodologia é eficaz e robusta, além de atingir com êxito o objetivo inicial.

Ademais, para que os resultados fiquem melhores do que os já obtidos, algumas alterações podem ser feitas como trabalho futuro, sendo elas: a utilização do áudio dos vídeos, pois a partir desse elemento chave, pode-se refinar bastante os resultados da rede, como feito em trabalhos do desafio [4].

Uma outra melhoria que pode ser feita, é otimizar os métodos utilizados, demarcando parâmetros específicos e definindo novas características para complementar as já existentes no trabalho, sendo necessário um aprofundamento para determinação de tais. Uma possível abordagem seria a utilização de técnicas para a extração de dados de descritores de Análise de Componentes Principais (PCA).

A partir desses aperfeiçoamentos, prevê-se uma melhoria na acurácia que poderá proporcionar uma futura evolução de reconhecimento para um detector automático de mentiras.

Referências

[1] Students, DH101 EPFL. `http://veniceatlas.epfl.ch/atlas/gis-and-databases/objects/recognizing-copies-of-paintings-1/`, Acessado em: 23 de Abril de 2017. ix, 9

[2] Haykin, Simon: *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edição, 1998, ISBN 0132733501. ix, 10, 11, 12, 19

[3] xbaro. `https://competitions.codalab.org/competitions/9181`, Acessado em: 10 de Abril de 2016. ix, 16, 31

[4] Ponce-López, Víctor, Baiyu Chen, Marc Oliu, Ciprian Cornearu, Albert Clapés, Isabelle Guyon, Xavier Baró, Hugo Jair Escalante e Sergio Escalera: *Chalearn lap 2016: First round challenge on first impressions - dataset and results*. 2016. 1, 2, 31, 33

[5] Ambady, Nalini, Frank J. Bernieri e Jennifer A. Richeson: *Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream*, volume 32 de *Advances in Experimental Social Psychology*, páginas 201–271. 2000, ISBN 0120152320. 1

[6] Hassin, Ran e Yaacov Trope: *Facing faces: Studies on the cognitive aspects of physiognomy*. Journal of Personality and Social Psychology, 78(5):837–852, maio 2000, ISSN 0022-3514. 1

[7] Berry, D.S.: *Taking people at face value: Evidence for the kernel of truth hypothesis. social cognition*. 8(4):343, 1990. 1

[8] Ambady, Nalini e Robert Rosenthal: *Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis*. março 1992. 1

[9] Willis J., Todorov A.: *First impressions: making up your mind after a 100-ms exposure to a face.* 1

[10] Ozer, Daniel J. e Verónica Benet-Martínez: *Personality and the prediction of consequential outcomes*. Annual Review of Psychology, 57(1):401–421, 2006. 2

[11] Roberts, B.W., Kuncel N.R. Shiner R. Caspi A. Goldberg L.R.: *The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes.* 2(4):313–345, 2007. 2

[12] Huffcutt, A.I., Conway J.M. Roth P.L. Stone N.J.: *Identification and meta-analytic assessment of psychological constructs measured in employment interviews.* 86(5):897, 2001. 2

[13] Nishant Hariprasad Pandey, Dr. Preeti Bajaj, Shubhangi Giripunje: *Design and implementation of real time facial emotion recognition system.* 2013. 2

[14] Wilson, Phillip Ian e John Fernandez: *Facial feature detection using haar classifiers.* J. Comput. Sci. Coll., 21(4):127–133, abril 2006, ISSN 1937-4771. `http://dl.acm.org/citation.cfm?id=1127389.1127416`. 2

[15] Pentland, M. Turk; A.: *Face recognition using eigenfaces.* página 586–591, 1991. 2

[16] Wei, X.S., Luo J.H. Wu J.: *Selective convolutional descriptor aggregation for fine-grained image retrieval.* 2016. 3

[17] LeCun, Y., Bottou L. Bengio Y. Haffner P.: *Gradient-based learning applied to document recognition.* 1998. 3

[18] Parkhi, O.M., Vedaldi A. Zisserman A.: *Deep face recognition.* 2015. 3

[19] Hochreiter, Sepp; e Jürgen Schmidhuber: *Long short-term memory.* 1997. 3

[20] He, K., Zhang X. Ren S. Sun J.:: *Deep residual learning for image recognition.* 2015. 3

[21] Gonzalez, Rafael C. e Richard E. Woods: *Digital Image Processing (3rd Edition).* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006, ISBN 013168728X. 5

[22] Trucco, Emanuele e Alessandro Verri: *Introductory Techniques for 3-D Computer Vision.* Prentice Hall PTR, Upper Saddle River, NJ, USA, 1998, ISBN 0132611082. 5

[23] Horn, B. K. P. e B. G. Schunck: *Determining optical flow.* Artificial Intelligence, 17:185–203, 1981. 5, 6, 7, 16

[24] Barron, J. L., D. J. Fleet e S. S. Beauchemin: *Performance of optical flow techniques.* Int. J. Comput. Vision, 12(1):43–77, fevereiro 1994, ISSN 0920-5691. `http://dx.doi.org/10.1007/BF01420984`. 5

[25] Lucas, Bruce D. e Takeo Kanade: *An iterative image registration technique with an application to stereo vision.* Em *In IJCAI81*, páginas 674–679, 1981. 7

[26] Anandan, P.: *A computational framework and an algorithm for the measurement of visual motion.* International Journal of Computer Vision, 2:283–310, 1989. 7, 8

[27] Burt P.J., Yen C. e Xu X.: *Multiresolution flow through-motion analysis.* IEEE CVPR, Washington, páginas 246–252, 1983. 7

[28] Glazer F., Reynolds G. e Anandan P.: *Scene matching through hierarchichal correlation.* IEEE CVPR, Washington, páginas 432–441, 1983. 7

[29] Little J.J., Bulthoff H.H. e Poggio T.A.: *Parallel optical flow using local voting.* IEEE CVPR, Tampa, páginas 454–459, 1988. 7

[30] Little J.J., Verri A.: *Analysis of diferential and matching methods for optical flow.* IEEE Workshop on Visual Motion, Irvine CA, páginas 173–180, 1989. 7

[31] Beauchemin, S. S. e J. L. Barron: *The computation of optical flow.* ACM Comput. Surv., 27(3):433–466, setembro 1995, ISSN 0360-0300. `http://doi.acm.org/10.1145/212094.212141`. 8

[32] E.H., Adelson e Bergen J.R.: *The extraction of spatiotemporal energy in human and machine vision.* IEEE Workshop on Visual Motion, Charleston, páginas 151–156, 1986. 8

[33] Barman H., Haglund L., Knutsson H. e Granlund G.: *Estimation of velocity acceleration and disparity in time sequences.* IEEE Workshop on Visual Motion, Princeton, páginas 44–5, 1991. 8

[34] Bigun J., Granlund G. e Wiklund J.: *Multidimensional orientation estimation with applications to texture analysis and optical flow.* IEEE Trans. PAMI, páginas 775–790, 1991. 8

[35] L., Haglund: *Adaptive multidimensional filtering.* 1992. 8

[36] D.J., Heeger: *Optical flow using spatiotemporal filters.* Int. J. Comp. Vision 1, páginas 279–302, 1988. 8

[37] B., Jahne: *Image sequence analysis of complex physical objects: nonlinear small scale water surface waves.* IEEE ICCV London, páginas 191–200, 1987. 8

[38] E.H., Adelson e Bergen J.R.: *Spatiotemporal energy models for the perception of motion.* J. Opt. Soc. Am., páginas 284–299, 1985. 8

[39] D.J., Fleet: *Measurement of image velocity.* 1992. 8

[40] van, Santen J.P.H. e Sperling G.: *Elaborated reichardt detectors.* J. Opt. Soc. Am., páginas 300–321, 1985. 8

[41] A.B., Watson e Ahumada A.J.: *Model of human visual-motion sensing.* J. Opt. Soc. Am., páginas 322–342, 1985. 8

[42] E.P., Simoncelli: *Distributed representation and analysis of visual motion.* 1993. 8

[43] Dalal, Navneet e Bill Triggs: *Histograms of oriented gradients for human detection.* Em *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, páginas 886–893, Washington, DC, USA. IEEE Computer Society, ISBN 0-7695-2372-2. `http://dx.doi.org/10.1109/CVPR.2005.177`. 8, 9

[44] Mohan, A., C. Papageorgiou e T. Poggio: *Example-based object detection in images by components.* PAMI, páginas 349–361, apr 2001. 9

[45] Belongie, S., J. Malik e J. Puzicha: *Matching shapes*. The 8th ICCV, Vancouver, Canada, páginas 454–461, 2001. 9

[46] Lowe, D. G.: *Object recognition from local scale-invariant features*. Em *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, páginas 1150–1157 vol.2, 1999. 9

[47] Björck, Åke: *Numerical Methods for Least Squares Problems*. Siam Philadelphia, 1996, ISBN 0-89871-360-9. 12

[48] Digman, J. M.: *Personality Structure: Emergence of the Five-Factor Model*. Annual Review of Psychology, 41(1):417–440, 1990. `http://dx.doi.org/10.1146/annurev.ps.41.020190.002221`. 14

[49] Lourakis, Manolis IA: *A brief description of the levenberg-marquardt algorithm implemented by levmar*. Foundation of Research and Technology, 4(1), 2005. 19

[50] Bradski, G. *Dr. Dobb's Journal of Software Tools*, 2000. 24

[51] The MathWorks, Inc. `http://www.mathworks.com/help/nnet/ref/trainlm.html`, Acessado em: 23 de Abril de 2017. 24