

UNIVERSIDADE FEDERAL DO PARANÁ

DIANDRA AKEMI ALVES KUBO

RECONHECIMENTO AUTOMÁTICO DE EXPRESSÕES FACIAIS

CURITIBA
2016

DIANDRA AKEMI ALVES KUBO

RECONHECIMENTO AUTOMÁTICO DE EXPRESSÕES FACIAIS

Trabalho apresentado como requisito parcial para a obtenção do grau de Bacharel do curso de Informática Biomédica, Setor de Ciências Exatas da Universidade Federal do Paraná.

Orientador: Prof^a. Dra. Olga Regina Pereira Bellon

CURITIBA
2016

Aos meus pais.

AGRADECIMENTOS

À Deus.

À minha família.

À Professora Dtra Olga Bellon e o Professor Dtr Luciano Silva, por desde cedo terem me acolhido no IMAGO.

Aos demais professores e funcionários da UFPR que de um jeito ou de outro contribuíram e continuam a contribuir na vida de todos os alunos.

Aos professores do IFPR, principalmente o Professor Dtr Roberto Teixeira Alves, que foi essencial na minha escolha por Informática Biomédica.

A todos meus amigos de departamento, mas principalmente: ao Thiago, Lucas e Fabrício pelo apoio sem o qual esse trabalho não seria concluído.

Aos meus amigos do SIMEPAR: Camila, Jorge, Renan, Mariana, Alana e Tiago pelo companheirismo na rotina diária.

Aos meus amigos da vida: Manuel, Rafael, Gabriel, Paulo e Talita por simplesmente existirem.

“Whatever you value, be committed to it and let nothing distract you from this goal. The uncommitted life, like Plato’s unexamined life, is not worth living.”

—The Hesburgh Papers, 1979

RESUMO

Expressões faciais humanas são um dos mais importantes canais de comunicação, podendo ser usadas com confiança para entender os estados da mente. De fato, quando a comunicação verbal não está disponível ou não é possível, analisar a expressão facial de uma pessoa pode ser uma forma confiável de caracterizar estados emocionais. Empresas de Marketing tem investido em análise de expressões faciais como parte do esforço de medir reações e níveis de interesse de consumidores. Em contraste, empresas da área de saúde estão estudando essa tecnologia na busca de maneiras melhores de assistir aqueles que não podem expressar verbalmente suas necessidades, como crianças com autismo, que devido a um distúrbio neurológico tem uma comunicação verbal comprometida. Como resultado desse interesse, muitos algoritmos e métodos para reconhecimento de expressões faciais tem sido desenvolvidos. Nesse contexto, revisamos a literatura caracterizando diferentes classes de métodos e conduzindo testes em algoritmos distintos afim de conseguir resultados condizentes com a performance de cada um na base de dados BU-3DEF. O melhor resultado obtido foi utilizando um combinação de Eigenfaces e SVM como classificador.

Palavras-chave: Expressões faciais, eigenfaces, gabor filters, CNN, SVM

ABSTRACT

Human facial expressions are one of the most important channels of communications, and can be reliably used to understand the states of the mind. In fact, when verbal communication is not available or not possible, analyzing one's facial expression can be a reliable way to label emotional state. Marketing firms have invested in facial expression analysis as part of their efforts to measure reactions and levels of interest of customers; by contrast, healthcare companies are studying the technology as they search for better ways to assist those who cannot verbally express their needs. As a result of this interest, many facial expression recognition algorithms and methods have been developed. We review the literature, characterize the different classes of methods, and run experiments on algorithms from these different classes to get tangible results. We undertook an analysis and evaluation of these methods as a first step to chose which one is the best for an application. Tests were made on the BU-3DEF database. The best results were given by a combination of Eigenfaces and a SVM classifier

Key-words: Facial expression, eigenfaces, gabor filters, CNN, SVM

LISTA DE FIGURAS

FIGURA 1 – FILTROS RETANGULARES DO MÉTODO VIOLA-JONES	17
FIGURA 2 – CASCATA ATENCIONAL	18
FIGURA 3 – NEURÔNIO	19
FIGURA 4 – REDE NEURAL	20
FIGURA 5 – REDE NEURAL CONVOLUCIONAL	21
FIGURA 6 – TRIANGULAÇÃO DE DELAUNAY	23
FIGURA 7 – LANDMARKS	24
FIGURA 8 – HIPERPLANO NO SVM E VETORES DE SUPORTE	26
FIGURA 9 – QUATRO CLASSES DE ALGORITMOS	28
FIGURA 10– ACURÁCIA MÉDIA POR NÚMERO DE COMPONENTES - ÂNGU- LOS	36
FIGURA 11– ACURÁCIA MÉDIA POR NÚMERO DE COMPONENTES - DIS- TÂNCIAS	36
FIGURA 12– ACURÁCIA MÉDIA POR EMOÇÃO - GABOR	39
FIGURA 13– ACURÁCIA MÉDIA POR EMOÇÃO - EIGENFACES	39
FIGURA 14– ACURÁCIA MÉDIA POR EMOÇÃO - REDE NEURAL	40
FIGURA 15– ACURÁCIA MÉDIA POR EMOÇÃO - ÂNGULOS	40
FIGURA 16– ACURÁCIA MÉDIA POR EMOÇÃO - DISTÂNCIAS	40
FIGURA 17– ACURÁCIA MÉDIA POR EMOÇÃO - COMPARAÇÃO	41
FIGURA 18– PRIMEIRAS 7 EIGENFACES NA BASE BU-3DEF	42
FIGURA 19– DIFERENTES EXPRESSÕES APARENTEMENTE SEMELHANTES	43
FIGURA 20– TRISTEZA CLASSIFICADA COMO RAIVA	49
FIGURA 21– MEDO CLASSIFICADO COMO FELICIDADE	50

FIGURA 22– RAIVA CLASSIFICADA COMO TRISTEZA 51

LISTA DE TABELAS

TABELA 1	– ABORDAGENS UTILIZADAS NA LITERATURA	27
TABELA 2	– REVISÃO CRONOLÓGICA DA LITERATURA	32
TABELA 3	– MATRIZ DE CONFUSÃO GABOR FILTERS	37
TABELA 4	– MATRIZ DE CONFUSÃO EIGENFACES	37
TABELA 5	– MATRIZ DE CONFUSÃO REDE DE CONVOLUÇÃO	37
TABELA 6	– MATRIZ DE CONFUSÃO ÂNGULOS	38
TABELA 7	– MATRIZ DE CONFUSÃO DISTÂNCIAS	38

LISTA DE SIGLAS

AAM	Active Appearance Models
FACS	Facial Action Coding System
HMM	Hidden Markov Model
LDA	Linear Discriminant Analysis
PCA	Principal Component Analysis
SOM	Self-Organizing Maps
SVM	Support Vector Machine

SUMÁRIO

1 INTRODUÇÃO	13
1.1 JUSTIFICATIVA	13
1.2 OBJETIVOS	14
1.3 DISPOSIÇÃO DO TRABALHO	14
2 FUNDAMENTAÇÃO TEÓRICA	15
2.1 DETECÇÃO DE FACE	15
2.1.1 Método Viola-Jones	16
2.2 EXTRAÇÃO DE CARACTERÍSTICAS	18
2.2.1 Redes Neurais	19
2.2.2 Autofaces (Eigenfaces)	21
2.2.3 Informação Geométrica	22
2.2.4 Filtros de Gabor	23
2.3 CLASSIFICAÇÃO	24
2.3.1 Support Vector Machine (SVM)	25
3 REVISÃO BIBLIOGRÁFICA	27
3.1 TAXONOMIA	27
3.2 DIVISÃO PELA EXTRAÇÃO DE CARACTERÍSTICAS	27
3.2.1 Atributos	27
3.2.2 Template	29
4 METODOLOGIA	33
4.1 DETECÇÃO DE FACE	33
4.2 EXTRAÇÃO DE CARACTERÍSTICAS	33
4.3 CLASSIFICAÇÃO	34

4.4	BASE DE DADOS	34
5	RESULTADOS	35
5.1	MATRIZ DE CONFUSÃO	37
5.2	GRÁFICOS	39
6	CONCLUSÃO	42
6.1	LIMITAÇÕES	43
6.2	TRABALHOS FUTUROS	44
	REFERÊNCIAS	45
	APÊNDICE A – EXEMPLOS DE EXPRESSÕES FACIAIS SIMILARES	49

1 INTRODUÇÃO

1.1 JUSTIFICATIVA

Em 1872, Darwin estabeleceu o princípio dos hábitos associados à expressão humana, escrevendo que “certas ações complexas estão sob direto ou indireto serviço de certos estados da mente, em ordem de aliviar ou gratificar certas sensações, desejos, etc.” (DARWIN, 1872). Ao fazer este comentário, Darwin nos impele a usar expressões para o entendimento dos sentimentos internos humanos.

A expressão verbal (principal canal de comunicação) não é possível em algumas ocasiões e, em outras, não é a que mais reflete o que uma pessoa possa estar sentindo. TIAN; KANADE; COHN (2005) demonstraram que a tarefa de interpretar expressões faciais humanas tem várias aplicações nas mais diversas áreas, requerindo assim, extensiva pesquisa. PANTIC; ROTHKRANTZ (2000a) mostram a tendência de transformar expressões faciais reconhecidas automaticamente em uma das seis emoções básicas definidas em EKMAN; FRIESEN (1971).

Essa tarefa pode também auxiliar no cuidado de neonatos (SCHIAVENATO *et al.*, 2008; HAZELHOFF *et al.*, 2009) e de adultos (LITTLEWORT; BARTLETT; LEE, 2009), assistindo-os na identificação de expressões de dor. Além disso, é possível medir o engajamento de uma pessoa (WHITEHILL *et al.*, 2014) enquanto a mesma realiza uma tarefa, que frequentemente é usada por departamentos de marketing de grandes companhias para se obter uma resposta da reação de consumidores e nível de interesse dos mesmos à suas propagandas/publicidade (GARBAS *et al.*, 2013; YEASIN; BULLOT; SHARMA, 2006). Pode-se também usar o reconhecimento de expressões faciais automático em sistemas de cuidados médicos para o monitoramento de qualidade de vida (LEE; CHUN; PARK, 2008), acompanhando as emoções dos usuários ou mesmo em uma sistema em que a interação humano-computador se dá por meio da

emoção do usuário em ordem de simplesmente mudar uma música (ABDAT; MAAOUI; PRUSKI, 2011). Em adição, uma esfera de aplicações dessa área envolve a tentativa de auxiliar no cuidado de crianças com Autismo (GUHA *et al.*, 2015).

1.2 OBJETIVOS

Esse trabalho tem como objetivo analisar a literatura existente em reconhecimento automático de expressões faciais, encontrar padrões que caracterizem e dividam os métodos da área de reconhecimento de expressões faciais e comparar esses diferentes tipos de métodos.

1.3 DISPOSIÇÃO DO TRABALHO

Este trabalho está estruturado em seis capítulos gerais, iniciados pela introdução. Nos capítulos 2 e 3 os materiais e métodos necessários para o entendimento da metodologia escolhida são apresentados.

No capítulo 2 os passos necessário para o reconhecimento de expressões são explanados. No capítulo 3 vários trabalhos presentes na literatura são apresentados e caracterizados de acordo com sua natureza.

No capítulo 4 a metodologia escolhida para realizar experimentos é mostrada e seus resultados aparecem no capítulo 5. As considerações finais são apresentadas no capítulo 6.

2 FUNDAMENTAÇÃO TEÓRICA

A tarefa de classificar automaticamente expressões faciais humanas é composta por basicamente três passos: detecção da face, extração de características da face e obtenção da classe à qual a mesma pertence. Nesse capítulo uma explicação para cada um desses passos é dada, e também são mostrados alguns métodos de cada passo.

2.1 DETECÇÃO DE FACE

A definição dada por (YANG; KRIEGMAN; AHUJA, 2002) afirma que detecção de faces é a tarefa de determinar se existem ou não faces em uma imagem, e caso existam, o resultado dessa tarefa deve ser a localização e extensão de cada face na imagem.

Em 1992, SAMAL; IYLANGAR apresentam um levantamento em métodos de análise de expressões faciais, e também apresentam um primeiro rascunho dos grupos de características que poderiam diferenciar tais métodos (SAMAL; IYLANGAR, 1992). Os autores afirmam que, para a primeira tarefa que é a detecção da face, existem dois grandes conjuntos: os que tem uma abordagem holística e os que tem uma abordagem baseada em características. Na abordagem holística existem os métodos que detectam a face como uma unidade inteira, em oposição à abordagem baseada em características, em que atributos e características da face são detectados primeiramente. KUMAR; BINDU (2006), HJELMÅS; LOW (2001), FASELA; LUETTINB (2002) e PANTIC; ROTHKRANTZ (2000a) também concordam com essa divisão dos métodos. Sendo um dos primeiros trabalhos de levantamento de métodos na área, SAMAL; IYLANGAR (1992) afirmavam que o problema de reconhecimento de expressões faciais não era seriamente pesquisado, discutindo apenas alguns trabalhos feitos na área,

cuja abordagem por um ponto de vista muscular era a principal no passo de extração de características, sugerindo que os métodos de análise de expressões deveriam utilizar algum tipo de aprendizado no futuro.

Adicionalmente, (YANG; KRIEGMAN; AHUJA, 2002) declaram que existem, efetivamente, quatro classes principais de métodos para detecção de faces, adicionando as que se baseiam em aprendizado e as que se baseiam em aparência. Métodos baseados em aprendizado são os construídos em regras codificadas a partir de conhecimento humano sobre o que constitui uma face. E, por seqüência, métodos baseados em aparência usam modelos aprendidos por meio de um conjunto de imagens que capturam a volatilidade da aparência facial humana.

2.1.1 Método Viola-Jones

Um dos métodos mais populares em detecção de faces é o método proposto por VIOLA; JONES (2004). O método, que leva o nome dos autores, tem seu treinamento demorado porém, na fase de detecção é extremamente rápido. As ideias principais são: utilizar cálculo de imagem integral para avaliação de características, *boosting* para seleção de características e cascata atencional para rejeição rápida de janelas que não contêm face.

O cálculo da imagem integral consiste em atribuir à cada *pixel* a soma dos *pixels* acima, à esquerda e dele mesmo. A imagem integral é calculada em apenas uma passada pela imagem original e é utilizada para obter as *Haar features*. Uma janela deslizante de um tamanho previamente definido passa pela imagem e em cada janela são obtidos os valores para cada filtro retangular. Esses atributos são obtidos calculando a subtração da intensidade dos *pixels* que estão na parte clara dos *pixels* da parte escura de um filtro retangular, como os mostrados na Figura 1. Como a combinação dos possíveis filtros para cada janela dentro de uma imagem é extremamente grande, uma seleção dos melhores é necessária.

Para selecionar os melhores atributos, várias rodadas de treinamento são reali-

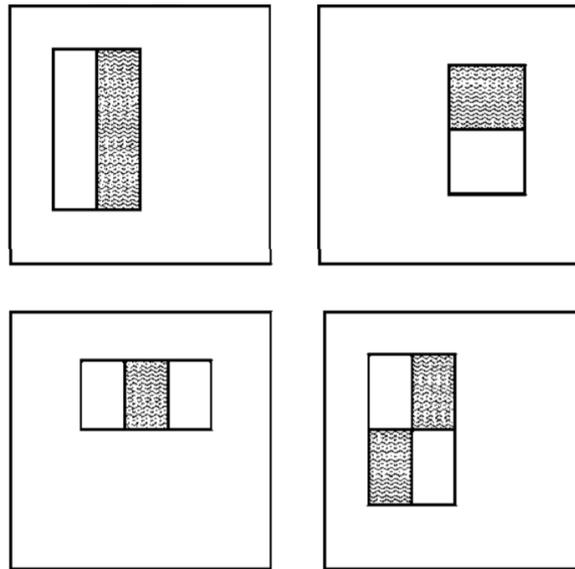


FIGURA 1: Filtros retangulares do método Viola-Jones

FONTE: (VIOLA; JONES, 2004)

zadas. Em cada rodada, um classificador fraco (um simples *threshold* nos valores obtidos pelo filtro retangular) que apresenta desempenho bom em exemplos que foram difíceis na rodada anterior, é escolhido. A informação de dificuldade é obtida por meio de pesos que são atribuídos a cada entrada do treinamento, em que as entradas classificadas erroneamente têm seu peso aumentado. O classificador final será, então, um conjunto de classificadores fracos que tem bom desempenho em algumas entradas, mas sua combinação linear resulta em um bom classificador.

A combinação se dá por meio de uma cascata atencional, em que, no nível mais alto, ficam os classificadores que mais rejeitam janelas sem faces e passam mais janelas com faces. A cada nível descido nessa cascata, mais complexo fica o classificador. Um esquema da cascata pode ser visto na Figura 2.

Após uma face ser detectada utilizando uma das abordagens discutidas, o próximo passo é extrair informação da mesma.

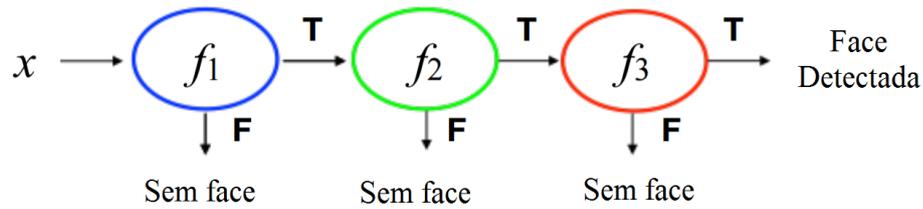


FIGURA 2: Cascata atencional

FONTE: A autora (2016)

2.2 EXTRAÇÃO DE CARACTERÍSTICAS

Extrair informação da face é crucial para a tarefa de classificação. Assim como PALI; GOSWAMI; BHAIYA (2014) afirmam: “Características são propriedades que descrevem a imagem inteira” e, em imagens faciais, o seu propósito é sumarizar a imagem mantendo suas propriedades originais.

Nessa etapa, (PANTIC; ROTHKRANTZ, 2000a) reivindicam que uma divisão pode ser feita entre algoritmos que usam *templates* (os que tentam combinar um modelo holístico à face detectada) e os que são baseados em atributos, que objetivam localizar ou seguir partes da face detectadas ao invés de tratá-la como apenas uma unidade.

FASELA; LUETTINB (2002) concordam com SAMAL; IYLANGAR (1992) e PANTIC; ROTHKRANTZ (2000a) em separar a detecção de faces entre os grupos holístico *versus* baseado em características. Porém, FASELA; LUETTINB (2002) vão além na definição de categorias no passo de extração de características. Inicialmente, fala-se sobre movimento *versus* deformação, em que, no último, é necessária a imagem da face em expressão neutra para que as mudanças sejam analisadas e, no primeiro, a detecção das mudanças deve ser feita desde o início, sem a expressão neutra como base. Adicionalmente, a segunda partição pode ser feita entre imagem *versus* abordagens baseadas em modelos, que é similar à classificação entre características *versus* *template* (PANTIC; ROTHKRANTZ, 2000a). Por fim, métodos baseados em aparências *versus* músculos são divididos: métodos que focam nos efeitos das atividades

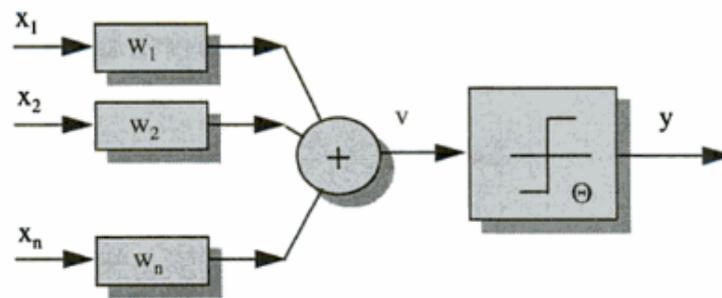


FIGURA 3: Neurônio

FONTE: KOVÁCS (2002)

musculares na face são colocados no primeiro grupo e os que tentam inferir atividade muscular a partir da informação visual, no segundo.

2.2.1 Redes Neurais

Em redes neurais, a intenção é mimetizar o processo classificatório do cérebro humano, simulando neurônios. A representação da simulação de um neurônio pode ser vista na Figura 3. No seu funcionamento mais simples, em que a classificação é binária (pertence ou não à classe), cada instância de dados a serem classificados tem o seu vetor de atributos $\{x_1, \dots, x_n\}$ multiplicado por um vetor de pesos $\{w_1, \dots, w_n\}$, que busca ter projeção positiva para exemplos positivos e projeção negativa para exemplos negativos. O somatório dessas multiplicações de vetores, somado a um bias w_0 ou b (também chamado de viés), passa por uma função de ativação ϕ , que produz a saída y . Essa função é responsável por determinar a intensidade dos valores que foram transmitidos, definindo se essa instância pertence ou não à classe inicial definida (FAUSETT, 1994).

Na rede, de modo geral, temos camadas de nós interconectados, em que cada nó produz o resultado de uma função não linear de seus dados de entrada (QUINLAN, 2014; MICHIE; SPIEGELHALTER; TAYLOR, 1994). Um exemplo de uma rede pode ser visto na Figura 4. Devido a essas conexões, um alto grau de não linearidade pode ser exigida da rede, dependendo da arquitetura escolhida para resolução do problema.

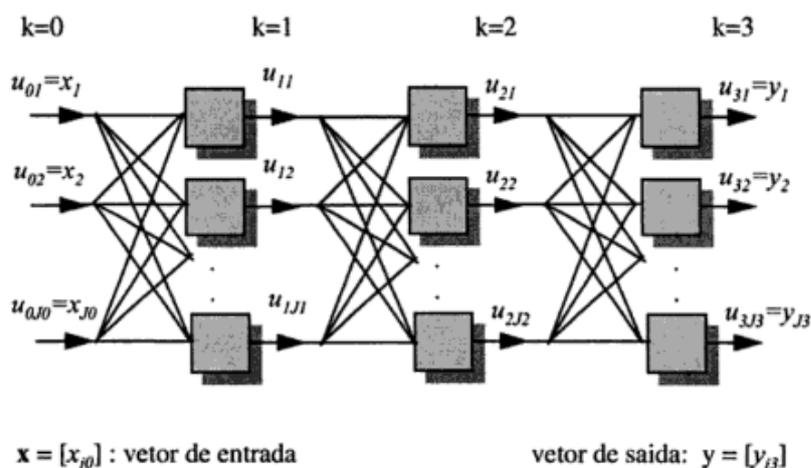


FIGURA 4: Rede Neural

FONTE: KOVÁCS (2002)

Após o dado passar pela rede inteira, eventualmente ele chega à um nó de saída, em que a classe final é calculada.

Uma variação de redes neurais são as redes neurais convolucionais, em que já se assume que as entradas serão imagens e, portanto, tem uma arquitetura e funções de ativações que permitem que certas propriedades de imagens sejam tratadas de maneira melhor (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Ao contrário de neurônios de redes normais, nas redes convolucionais os neurônios tem três dimensões: altura, largura e profundidade, o que é de extrema importância para o tratamento de imagens. Ao final da rede o resultado continuará sendo composto por um vetor de saída com os scores para cada classe.

As camadas de convolução, que são o que definem esse tipo de rede, são camadas cujos filtros das imagens são aprendidos por meio do algoritmo de *backpropagation* (HECHT-NIELSEN, 1989). Um exemplo de rede convolucional pode ser visto na Figura 5. Ao final da fase de treinamento, é por meio da situação final dessas camadas que a extração de características pode ser realizada.

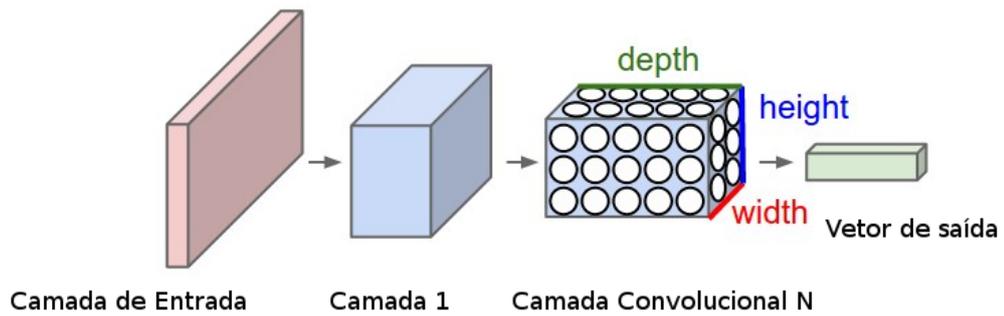


FIGURA 5: Rede Neural Convolutiva

2.2.2 Autofaces (Eigenfaces)

As Autofaces (Eigenfaces) (TURK; PENTLAND, 1991) de um conjunto de imagens podem ser obtidas realizando uma análise de componentes principais em cima das mesmas. Uma vez que todas as imagens a serem analisadas tem a mesma dimensão ($h \times w$), as mesmas são transformadas em vetores que compõem uma matriz M de dimensão ($n \times x$), em que n representa o número de imagens e $x = h \times w$.

Uma imagem média é calculada (Equação 1) e nada mais é do que uma imagem em que cada *pixel* representa a média dos valores dos *pixels* naquela mesma posição na base de imagens sendo analisada.

$$\mu = \left(\frac{1}{n}\right) \times \sum_{i=1}^n img_i \quad (1)$$

Cada imagem img_i tem a imagem média μ subtraída de si, para que as mudanças sejam analisadas (Equação 2).

$$new_img_i = img_i - \mu \quad (2)$$

A partir da matriz M , cujas linhas são cada new_img_i , a matriz de covariância

(Equação 3) é calculada. Dela, os autovalores e autovetores são obtidos. Nesse caso, os autovetores, quando transformados em matriz, também representam uma face e por isso são chamados de autofaces ou *eigenfaces*, que serão as características a serem analisadas.

$$C = M^T M \quad (3)$$

Os *eigenvectors* EV são ordenados de acordo com seus autovalores (*eigenvalues*), o número desejados deles é selecionado e então ocorre uma multiplicação pela matriz M , de acordo com a Equação 4, que resulta no espaço de faces FS .

$$FS = M \times EV \quad (4)$$

Finalmente, as imagens são projetadas multiplicando-as pelo espaço de faces transposto, e então os dados passam para o classificador escolhido (Equação 5).

$$P = FS^T \times M \quad (5)$$

Esse método foi originalmente projetado para utilização em reconhecimento facial (TURK; PENTLAND, 1991), que é a tarefa de identificar a quem pertence a face analisada. Porém, outros trabalhos também o utilizam para reconhecimento de expressões faciais (PADGETT; COTTRELL, 1997; WANG *et al.*, 2010).

2.2.3 Informação Geométrica

Extrair a informação geométrica da face significa obter as formas e localizações de componentes faciais (TIAN; KANADE; COHN, 2005). Uma forma de utilizar essa informação pode ser vista em BENITEZ-QUIROZ; SRINIVASAN; MARTINEZ (2016),

em que os *landmarks* faciais são detectados e, a partir deles, uma triangulação de *Delaunay* (LEE; SCHACHTER, 1980) é realizada. Assim, obtêm-se ângulos internos desses triângulos cujos vértice são os próprios *landmarks*. Dessa forma, pode-se inferir a informação de deformação do rosto, que pode ser importante na tarefa de classificação da expressão facial. Um exemplo da triangulação de *Delaunay* pode ser observado na imagem 6.

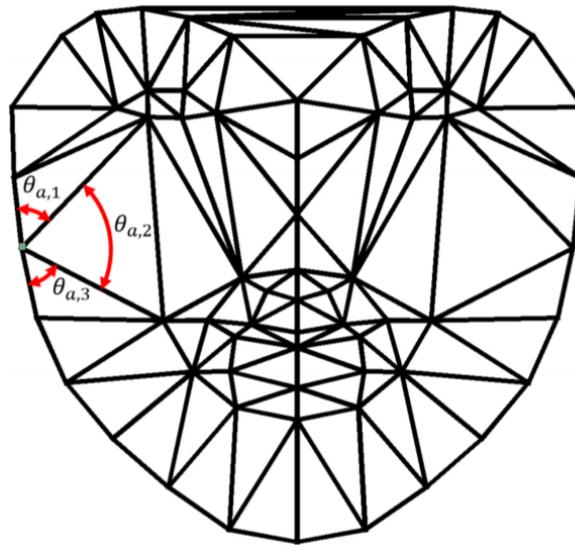


FIGURA 6: Triangulação de Delaunay

FONTE: BENITEZ-QUIROZ; SRINIVASAN; MARTINEZ (2016)

$$distancia = \sqrt{(x1 - x2)^2 + (y1 - y2)^2} \quad (6)$$

Outro exemplo de características que se pode obter são as distâncias entre os pontos faciais detectados, como a distância euclidiana, que pode ser observada na Equação 6. Exemplo de landmarks faciais podem ser vistos na Figura 7.

2.2.4 Filtros de Gabor

Essa característica pode ser extraída aplicando uma máscara em cima dos valores da imagem original. Os valores de uma máscara são definidos por um função, que,



FIGURA 7: Landmarks

FONTE: A autora (2016)

nesse caso, pode ser encontrada na Equação 7 (GRIGORESCU; PETKOV; KRUIZINGA, 2002).

$$g(x, y; \gamma, \theta, \psi, \sigma, \lambda) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x'}{\lambda} + \psi\right) \quad (7)$$

O theta (θ) define a direção do filtro. O sigma (σ) é o desvio padrão da função gaussiana utilizada. Já o lambda (λ) é o tamanho de onda do fator sinusoidal. O gamma (γ) é a relação do aspecto espacial. E o psi (ψ) é o offset da fase.

2.3 CLASSIFICAÇÃO

Essa tarefa consiste em transformar os dados extraídos da face em informação sobre sua expressão facial. As classes finais podem ser representativas das expres-

sões faciais básicas, similares às que são descritas no Sistema de Codificação de Ação Facial (*Facial Action Coding System - FACS*) (EKMAN; ROSENBERG, 1997), ou podem ser instâncias de emoções (EKMAN; FRIESEN, 1971), que também são frequentemente associadas às expressões faciais.

PANTIC; ROTHKRANTZ (2000a) apresentam uma última partição de métodos por sua maneira de classificar expressões faciais, alocando-os em três grupos: baseados em modelos, em regras ou baseados em redes neurais artificiais. Os métodos baseados em modelos são os que, a partir do conjunto de características extraído, de alguma maneira, realizam uma comparação com modelos que representam cada expressão, possibilitando uma análise de similaridades que, conseqüentemente resulta na classificação. Métodos baseados em regras usam um conjunto de regras que define cada expressão facial e, quando essa regra é realizada, uma expressão facial é encontrada. Por fim, a última categoria refere-se aos métodos que utilizam redes neurais artificiais previamente treinadas como classificador.

Muitas distinções podem ser feitas entre os trabalhos realizados na área de classificação de expressões faciais, o que resulta num agrupamento dos mesmos em diferentes grupos. A natureza da implementação de cada um dos passos discutidos nessa seção é o que define as características do algoritmo final, conseqüentemente, definindo sua categoria.

2.3.1 Support Vector Machine (SVM)

Assim como (MICHIE; SPIEGELHALTER; TAYLOR, 1994) afirmam, para que a classificação de um novo contexto possa ser feita, é necessário que, previamente, algum tipo de processo ou aprendizado de contextos conhecidos seja realizada. No método SVM (*Support Vector Machine*), o objetivo é construir um hiperplano que divida ao máximo os exemplos sendo aprendidos.

Quando esses exemplos não são linearmente separáveis, o método busca uma função de mapeamento que projete os exemplos em um novo espaço de dados em

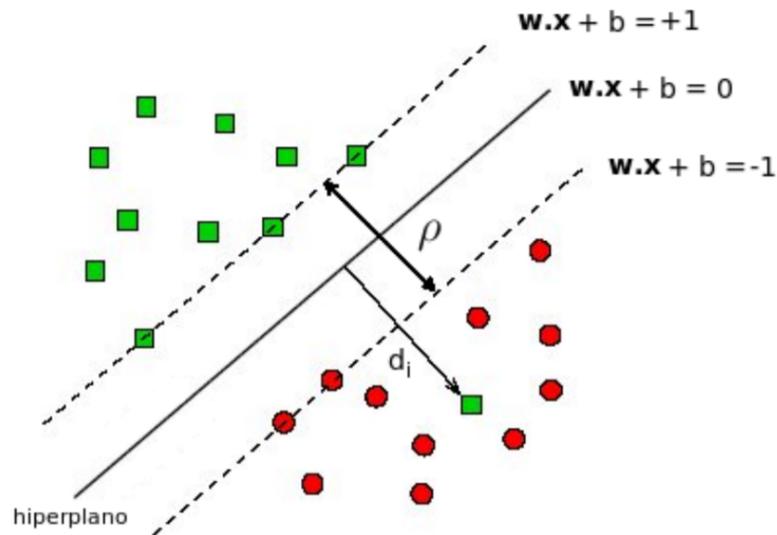


FIGURA 8: Hiperplano no SVM e vetores de suporte

FONTE: GONCALVES (2014)

que os mesmos sejam separáveis (GONCALVES, 2014). O hiperplano ótimo será o que mantém uma distância igual à todas as classes. Na figura 8, pode-se ver um hiperplano ótimo definido pelos vetores de suporte, os exemplos da base de treinamento mais próximos.

Muitas distinções podem ser feitas entre os trabalhos realizados na área de classificação de expressões faciais, o que resulta num agrupamento dos mesmos em diferentes grupos. A natureza da implementação de cada um dos passos discutidos nesse capítulo é o que define as características do algoritmo final, consequentemente, definindo sua categoria.

3 REVISÃO BIBLIOGRÁFICA

3.1 TAXONOMIA

Existem diversas maneiras de caracterizar um classificador de Expressões Faciais, como foi mostrado no capítulo 2. Na Tabela 1 todas as categorias são mostradas, juntamente os passos em que elas estão relacionadas.

Abordagem	Deteção de Face	Extração de Caract.	Classificação
Características	X	X	
Holística	X	X	
Conhecimento	X	X	X
Aparência	X	X	
Movimento		X	
Deformação		X	
Músculo		X	
Template		X	X
Redes Neurais	X	X	X

TABELA 1: Abordagens utilizadas na literatura

FONTE: A autora (2016)

3.2 DIVISÃO PELA EXTRAÇÃO DE CARACTERÍSTICAS

Em ordem de melhor compreender o trabalho realizado nessa área, os trabalhos revisados são divididos em dois grandes conjuntos: *Atributos* e *Template*. Cada um deles se interlaça com os conjuntos de Movimento e Deformação, como pode ser visto na Figura 9.

3.2.1 Atributos

Assim como foi mostrado na Seção 2, algoritmos que focam nos atributos da face têm uma abordagem local às imagens que serão analisadas em ordem de extrair da-

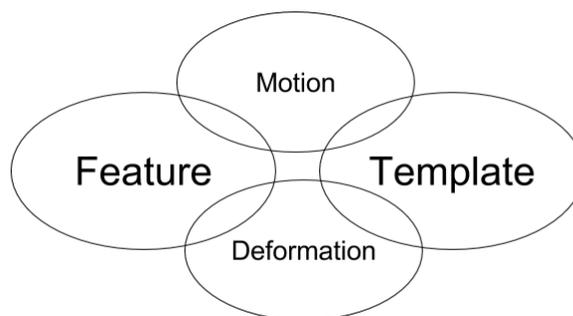


FIGURA 9: Quatro classes de algoritmos

FONTE: A autora (2016)

dos da face.

PILOWSKY; KATSIKITIS (1994) utilizaram o programa de taxonomia numérica *SNOB* para verificar se o algoritmo particionaria uma base de dados de expressões faciais nas seis expressões de emoções fundamentais de Ekman (EKMAN; FRIESEN, 1971). A entrada para o programa continha 62 pontos detectados na face. O método encontrou 5 classes, sendo que uma representa expressão de felicidade, outra a total ausência de expressão de felicidade e, as outras três com as demais expressões variadas, mostrando a dificuldade na distinção entre essas expressões.

KATOH; FUKUI (1998) empregaram o uso de Mapas Auto Organizáveis para extrair características de sequencias de imagens, em que um mapa para cada característica foi treinado (sobrancelha, olhos e boca), tendo os níveis de cinza de cada segmento como entrada. Posteriormente, a análise do neurônio vencedor para cada mapa mostrou uma consistência com as seis emoções de Ekman (KATOH; FUKUI, 1998).

LITTLEWORT *et al.* (2003) utilizaram detectores de características treinados com técnicas de *boosting* e o reconhecimento de expressões foi realizado com uma combinação de *AdaBoost* e *Support Vector Machine* (SVM).

Ao analisar as características extraídas do rosto, a abordagem de deformação compara a situação atual e a situação do rosto com a expressão neutra.

Em SUWA; SUGIE; FUJIMORA (1978) um dos primeiros trabalhos na área de ex-

pressões faciais é mostrado. Primeiramente, pontos foram desenhados no rosto da pessoa que teria sua foto tirada. A distância entre determinados pontos é calculada e transformada para um valor que quantifica a mudança para a expressão neutra. A classificação é feita utilizando a distância euclidiana entre os parâmetros que quantificam as mudanças.

TERZOPOULOS; WATERS em 1990 propuseram um sistema que era capaz de analisar sequencias de imagens e estimar as ações musculares que estavam ocorrendo nas mesmas, sendo assim capaz de reproduzir a mesma expressão em um modelo 3D, utilizando apenas a informação de localização de certos pontos faciais e a informação da expressão neutra.

Quando o que se verifica é o movimento, o algoritmo não leva em consideração a expressão neutra, que não existe em boa parte das bases de dados. Ao invés, analisa continuamente o estado atual sem o relacionar à expressão neutra.

MOSES; REYNARD; BLAKE (1995) reportam que o uso apenas da boca como parâmetro a ser analisado para classificar a expressão facial é suficiente. Nesse trabalho, o contorno da boca é rastreado em tempo real e a expressão é classificada por um classificador Bayesiano.

LIEN *et al.* (1998) trata a classificação de expressões baseada nas unidades de ação do Sistema de Codificação de Ação Facial (FACS). A discriminação foi desempenhada utilizando modelos de Markov (HMM) e a extração, por rastreamento de pontos e *dense flow*, com o auxílio da análise de componentes principais (PCA) e detecção de componentes de alto gradiente.

3.2.2 Template

A abordagem por extração de características por *template* consiste em considerar a face como um todo ao invés de analisá-la por partes.

BARTLETT *et al.* (2005) fizeram uma comparação de desempenho entre *AdaBo-*

ost, SVM e LDA (Análise Discriminante Linear) como classificadores. Os atributos são representados por filtros de *Gabor* da face. Os melhores resultados foram alcançados por uma combinação de *AdaBoost* para selecionar os atributos e uma cascata de SVMs como classificadores.

SAATCI; TOWN (2006) utilizaram *Active Appearance Models* para extrair o estado da face e uma cascata de SVMs para classificar os dados.

Para identificar até mesmo as características mais discriminativas das expressões mais sutis, SU; SATO (2013) utilizaram classificadores fracos para aprender as variações temporais dos atributos dessas expressões sutis de acordo com a ordem temporal dos mesmos. Logo após, um método de propagação de pesos foi aplicado para melhorar o desempenho dos classificadores fracos que fazem o reconhecimento precoce.

GARG; BAJAJ (2015) realizaram a extração de características utilizando o método de análise de componentes independentes, seguido por uma otimização de atributos utilizando um algoritmo genético e classificação utilizando uma rede neural artificial com *back propagation*.

Assim como na abordagem por características, na abordagem por *template* também é possível estudar a deformação e movimento da face.

ULUKAYA; ERDEM (2012) abordaram o problema de quantificar a deformação da face primeiramente estimando a expressão neutra da pessoa cuja expressão se tenta classificar. Um modelo Gaussiano é aplicado para obter a estimativa da expressão neutra e o que passa a ser analisado é a diferença entre a imagem analisada e a expressão neutra artificial. A classificação é realizada utilizando SVM.

CHETTY (2008) propôs um método em que, a partir da expressão neutra, a intensidade da expressão é calculada e uma função de Diferenciação Volumétrica por Setor é utilizada na classificação.

LYONS; BUDYNEK; AKAMATSU (1999) aplicaram uma abordagem com corres-

pondência de grafos elásticos rotulados com *gabor-wavelets* e extração de características pelo algoritmo de *eigenfaces*. A classificação foi, então, realizada pelo método de análise discriminante linear (*LDA*).

O trabalho de COHEN *et al.* (2003) foca na verificação de desempenho em imagens estáticas e dinâmicas. Nas imagens estáticas, um rastreador de deformação de volumes de Bezier foi utilizado para extrair características e uma rede Bayesiana para classificar os dados. Nas imagens dinâmicas foi utilizado um modelo escondido de Markov para extrair características e também para classificar.

No trabalho de PARK; KIM (2008) recorreu-se ao uso de magnificação de movimento por uma combinação de *Active Appearance Models* e vetores de movimento. O classificador escolhido foi o *SVM*.

Na tabela 2 os trabalhos podem ser observados em ordem cronológica.

Trabalho	Ano	Detec. de Face	Extração de Características	Classificação
(TERZOPOULOS; WATERS, 1990)	1990	Característica	Deformação	-
(PILOWSKY; KATSIKITIS, 1994)	1994	Característica	Característica	-
(MOSES; REYNARD; BLAKE, 1995)	1995	Característica	Característica e Movimento	Template
(LIEN <i>et al.</i> , 1998)	1998	Característica	Característica e Movimento	Template e Conhecimento
(KATOH; FUKUI, 1998)	1998	Característica	Característica	Conhecimento
(LYONS; BUDYNEK; AKAMATSU, 1999)	1999	Híbrida	Template	Template
(PANTIC; ROTHKRANTZ, 2000b)	2000	Híbrida	Híbrida	Conhecimento
(LITTLEWORT <i>et al.</i> , 2003)	2003	Holística	Característica	Template
(COHEN <i>et al.</i> , 2003)	2003	Híbrida	Template e Movimento	Template
(BARTLETT <i>et al.</i> , 2005)	2005	Híbrida	Template	Template
(SAATCI; TOWN, 2006)	2006	Holística	Template	Template
(PARK; KIM, 2008)	2008	Holística	Template e Movimento	Template
(CHETTY, 2008)	2008	Holística	Template e Deformação	Template
(ULUKAYA; ERDEM, 2012)	2012	Holística	Template e Deformação	Template
(SU; SATO, 2013)	2013	Holística	Híbrida	Template
(GARG; BAJAJ, 2015)	2015	Holística	Template	Template

TABELA 2: Revisão Cronológica da Literatura

FONTE: A autora (2016)

4 METODOLOGIA

4.1 DETECÇÃO DE FACE

A detecção da face foi realizada utilizando o método Viola-Jones explicado no capítulo 2.

4.2 EXTRAÇÃO DE CARACTERÍSTICAS

Como o foco deste trabalho é a comparação das características, cada uma foi testada separadamente para se ter uma ideia do desempenho natural.

A primeira rodada de testes foi em cima dos filtros de Gabor. De acordo com o que foi explicado no Capítulo 2, o tamanho do kernel/máscara utilizado foi de 30×30 . O theta (θ), que define a direção variou de 0 a π com um passo de $\frac{\pi}{8}$. O sigma (σ) foi definido como 4.0, que é o desvio padrão da função gaussiana utilizada. Já o lambda (λ) levou o valor de 10.0, que é o tamanho de onda do fator sinusoidal. O gamma (γ) teve valor 0.5, que é a relação do aspecto espacial. E o psi (ψ) valor 0.0, que é o offset da fase.

O segundo conjunto de categorias foram as Eigenfaces, onde as 150 primeiras eigenfaces extraídas a partir da matriz de covariância foram consideradas.

O terceiro conjunto foram as características extraídas por meio de uma rede neural de convolução, e a arquitetura escolhida foi a definida por LEVI; HASSNER em (LEVI; HASSNER, 2015).

Quanto à informação geométrica, primeiramente foram utilizados os ângulos obtidos na triangulação de Delaunay, e posteriormente as distâncias euclidianas entre os pontos do rosto.

Na triangulação, os 68 pontos encontrados na face resultaram em 114 triângulos, de onde conseqüentemente derivam-se 342 ângulos.

4.3 CLASSIFICAÇÃO

Por ser um dos classificadores mais utilizados na literatura, o método *SVM - Support Vector Machine* (explicado na Seção 2) foi escolhido para testar os diferentes atributos.

Todos os testes realizados foram feitos a partir de uma partição de 25% dos dados para teste e 75% para treinamento do método SVM.

4.4 BASE DE DADOS

A base de dados de expressões faciais utilizada foi a *BU-3DEF* da Universidade de Binghamton.

Essa base contém imagens de 100 participantes, 56 mulheres e 44 homens. A idade dos participantes varia entre 18 e 70 anos. Cada participante performou as seis expressões faciais básicas (felicidade, tristeza, surpresa, raiva, nojo, medo) em adição à expressão neutra (ZHANG *et al.*, 2013, 2014).

5 RESULTADOS

Na análise de desempenho inicial, o vetor de distâncias e de ângulos não apresentaram acurácia satisfatória. Por serem duas características com muita presença na literatura, foi decidido fazer uma análise de componentes principais para verificar se o mesmo afetaria no desempenho. A partir dessa análise, uma seleção (e transformação) de atributos por meio do método PCA (WOLD; ESBENSEN; GELADI, 1987). Um gráfico contendo o desempenho para cada número de componentes pode ser observado na Figura 10 para o vetor de ângulos, e na Figura 11 para o vetor de distâncias.

No caso do vetor de ângulos, o melhor desempenho foi alcançado com 200 componentes.

O vetor de distâncias inicialmente tinha tamanho 2278, pois é composto pela distância de todas as combinações não repetidas dos 68 pontos par a par (Equação 8). Após PCA foi transformado em um vetor de tamanho 10, o que é consistente com a literatura em expressões faciais que geralmente usa apenas algumas distâncias para medir a expressão facial (SCHIAVENATO *et al.*, 2008). Os resultados mostrados a seguir para essas duas características são então, resultado desse análise de componentes principais.

$${}_{(68)}\text{Combinacoes}_{(2)} = \frac{68!}{2! \times (68 - 2)!} \quad (8)$$

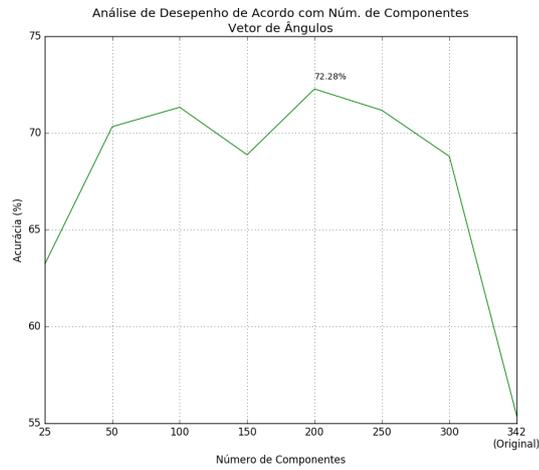


FIGURA 10: Acurácia Média por Número de Componentes - Ângulos
FONTE: A autora (2016)

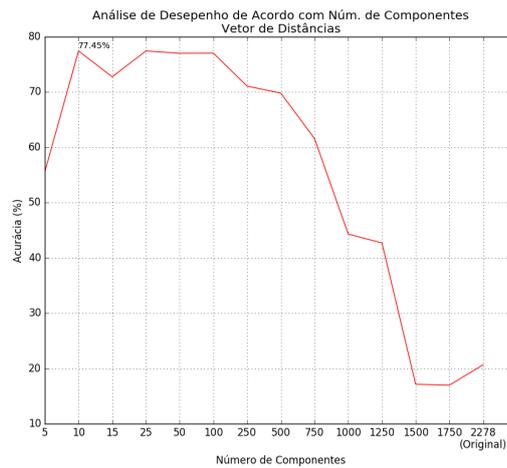


FIGURA 11: Acurácia Média por Número de Componentes - Distâncias
FONTE: A autora (2016)

5.1 MATRIZ DE CONFUSÃO

A matriz de confusão com a porcentagem de exemplos em cada categoria pode ser vista nas Tabelas 3 a 7.

	Nojo	Surpresa	Raiva	Medo	Tristeza	Felicidade
Nojo	69.1	4.5	12.7	7.3	2.7	3.6
Surpresa	6.5	81.7	2.2	4.3	2.2	3.2
Raiva	9.2	1.0	69.4	2.0	14.3	4.1
Medo	10.8	8.8	7.8	55.9	5.9	10.8
Tristeza	9.3	5.2	19.6	7.2	57.7	1.0
Felicidade	5.9	5.0	3.0	10.9	0.0	75.2

TABELA 3: Matriz de Confusão Gabor Filters

	Nojo	Surpresa	Raiva	Medo	Tristeza	Felicidade
Nojo	67.1	1.3	11.4	12.7	6.3	1.3
Surpresa	1.3	93.5	1.3	0.0	3.9	0.0
Raiva	6.4	0.0	83.3	1.3	6.4	2.6
Medo	3.8	6.2	3.8	76.2	1.2	8.8
Tristeza	6.2	1.2	21.2	7.5	63.7	0.0
Felicidade	1.2	2.5	2.5	10.0	1.2	82.5

TABELA 4: Matriz de Confusão Eigenfaces

	Nojo	Surpresa	Raiva	Medo	Tristeza	Felicidade
Nojo	71.4	0.8	10.1	10.9	4.2	2.5
Surpresa	0.8	80.8	2.5	7.5	5.8	2.5
Raiva	7.6	0.8	70.3	2.5	16.9	1.7
Medo	7.5	9.2	5.8	54.2	9.2	14.2
Tristeza	1.7	1.7	16.8	9.2	68.9	1.7
Felicidade	1.7	4.2	2.5	9.2	0.8	81.7

TABELA 5: Matriz de Confusão Rede de Convolução

	Nojo	Surpresa	Raiva	Medo	Tristeza	Felicidade
Nojo	78.2	0.0	11.8	4.5	2.7	2.7
Surpresa	2.2	91.4	2.2	0.0	0.0	4.3
Raiva	9.2	0.0	68.4	1.0	21.4	0.0
Medo	11.7	5.8	7.8	58.3	6.8	9.7
Tristeza	7.3	2.1	29.2	4.2	57.3	0.0
Felicidade	4.0	3.0	5.0	6.9	1.0	80.2

TABELA 6: Matriz de Confusão Ângulos

	Nojo	Surpresa	Raiva	Medo	Tristeza	Felicidade
Nojo	78.2	0.0	11.8	4.5	2.7	2.7
Surpresa	2.2	91.4	2.2	0.0	0.0	4.3
Raiva	9.2	0.0	68.4	1.0	21.4	0.0
Medo	11.7	5.8	7.8	58.3	6.8	9.7
Tristeza	7.3	2.1	29.2	4.2	57.3	0.0
Felicidade	4.0	3.0	5.0	6.9	1.0	80.2

TABELA 7: Matriz de Confusão Distâncias

5.2 GRÁFICOS

Os gráficos com as acurácias médias por emoção podem ser observados nas Figuras 12, 13, 14, 15 e 16.

Finalmente, na Figura 17 um gráfico com a acurácia média geral de todos os métodos testados é mostrada.

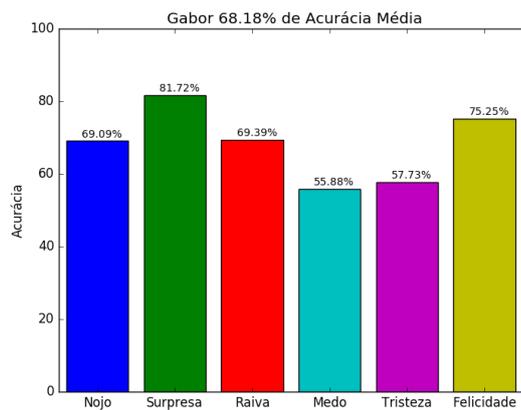


FIGURA 12: Acurácia Média por Emoção - Gabor

FONTE: A autora (2016)

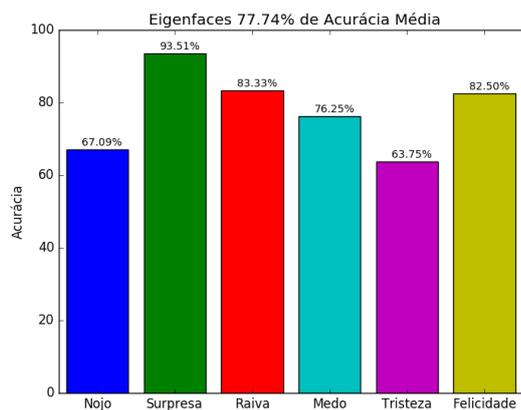


FIGURA 13: Acurácia Média por Emoção - Eigenfaces

FONTE: A autora (2016)

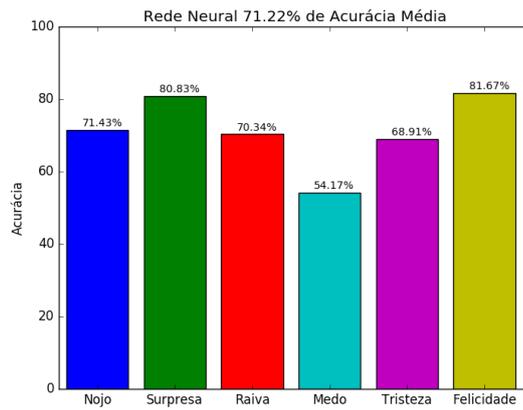


FIGURA 14: Acurácia Média por Emoção - Rede Neural

FONTE: A autora (2016)

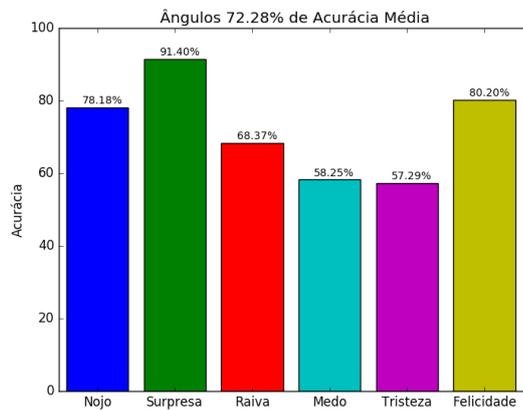


FIGURA 15: Acurácia Média por Emoção - Ângulos

FONTE: A autora (2016)

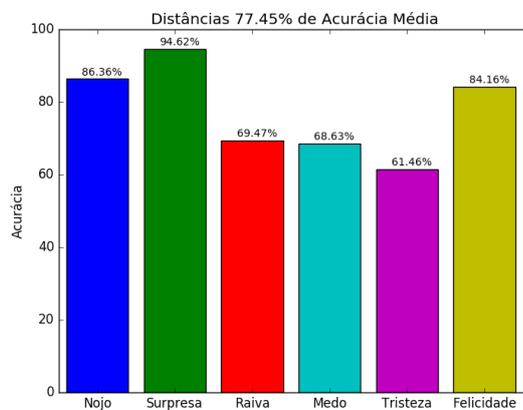


FIGURA 16: Acurácia Média por Emoção - Distâncias

FONTE: A autora (2016)

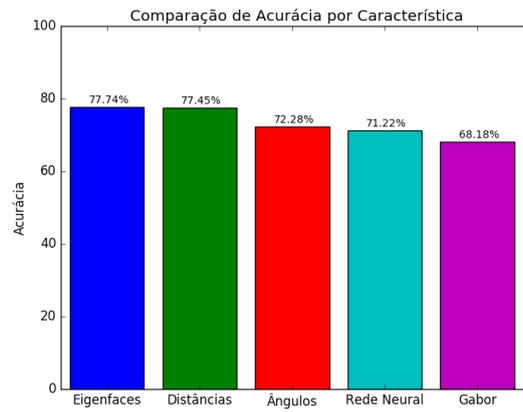


FIGURA 17: Acurácia Média por Emoção - Comparação

FONTE: A autora (2016)

6 CONCLUSÃO

Analisando os gráficos e as matrizes de confusão do desempenho das características analisadas, observamos que o algoritmo SVM combinado com as eigenfaces obteve o melhor resultado.

Na figura 18 podemos observar as 7 primeiras eigenfaces encontradas pelo algoritmo da PCA.



FIGURA 18: Primeiras 7 Eigenfaces na base BU-3DEF

FONTE: A autora (2016)

Apesar de ter sido a melhor característica, a combinação de Eigenfaces com SVM produziu uma acurácia de 77.75%. Para uma característica sozinha esse é um resultado relativamente bom, visto que os trabalhos mais recentes com resultados bons investem em *deep learning* (métodos que produzem vetores de características potencialmente maiores) ou algoritmos de *bagging* ou *boosting* que visam a combinação de características e classificadores. Como o objetivo da comparação desse trabalho era apenas comparar o desempenho das características sozinhas, esse resultado era esperado.

Esse problema ainda é um problema intrinsecamente difícil. Na Figura 19 isso pode ser observado. Na fileira de cima estão imagens classificadas como Tristeza, quando a expressão pedida era Raiva. E na fileira de baixa temos o contrário: expressões com a intenção de representar Tristeza classificadas como Raiva.



FIGURA 19: Diferentes expressões aparentemente semelhantes

FONTE: A autora (2016)

Ao analisar as imagens, nota-se uma clara semelhança entre as mesmas, e uma visível dificuldade na tarefa de torná-las separáveis. No Apêndice A mais exemplos podem ser visualizados.

6.1 LIMITAÇÕES

A área de reconhecimento de expressão facial tem muitas limitações devido ao fato de que a maioria das bases de dados são compostas por expressões posadas, o que implica em duas situações: pessoas diferentes não expressam emoções da mesma maneira e expressões posadas diferem de expressões autênticas.

A primeira situação é influenciada pelo fato de que diferentes culturas podem expressar a mesma emoção de maneira diferente, e dissemelhantes experiências e influências também moldam diferentes respostas à certas situações.

Na segunda situação, temos o dilema da base de treinamento seguir uma distribuição diferente da base de teste. Por um lado, se os dados de treinamento são provenientes de uma base com expressões “artificiais”, eles podem não ser úteis em uma aplicação *in the wild*, na vida real. Se por outro lado a base de teste também é proveniente de um ambiente de coleta não controlado, não existe um método que seja

totalmente confiável para saber qual é a expressão verdadeira que a pessoa capturada tinha intenção de fazer/sentir.

Além disso, essa área também sofre dos principais problemas que afetam a área de detecção de face, primeiro passo do fluxograma de algoritmos na área.

6.2 TRABALHOS FUTUROS

Portanto, é natural que o curso para trabalhos futuros sigam a direção do estudo aprofundado de métodos que sejam invariantes à pose da face, à oclusão parcial e que funcionem para dados coletados em situações reais e não encenadas.

REFERÊNCIAS

- ABDAT, F.; MAAOUI, C.; PRUSKI, A. Human-computer interaction using emotion recognition from facial expression. In: IEEE. **Computer Modeling and Simulation (EMS), 2011 Fifth UKSim European Symposium on**. [S.l.], 2011. p. 196–201.
- BARTLETT, M. S. *et al.* Recognizing facial expression: machine learning and application to spontaneous behavior. **Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on**, v. 2, p. 568–573, 2005.
- BENITEZ-QUIROZ, C. F.; SRINIVASAN, R.; MARTINEZ, A. M. Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In: **Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR'16), Las Vegas, NV, USA**. [S.l.: s.n.], 2016.
- CHETTY, G. A novel shape transformation approach for quantizing facial expressions. In: IEEE. **Digital Image Computing: Techniques and Applications (DICTA), 2008**. [S.l.], 2008. p. 168–175.
- COHEN, I. *et al.* Facial expression recognition from video sequences: temporal and static modeling. **Computer Vision and image understanding**, Elsevier, v. 91, n. 1, p. 160–187, 2003.
- DARWIN, C. **The expression of the emotions in man and animals**. [S.l.]: Oxford University Press, 1872.
- EKMAN, P.; FRIESEN, W. V. Constants across cultures in the face and emotion. **Journal of personality and social psychology**, American Psychological Association, v. 17, n. 2, p. 124, 1971.
- EKMAN, P.; ROSENBERG, E. L. **What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)**. [S.l.]: Oxford University Press, 1997.
- FASELA, B.; LUETTINB, J. Automatic facial expression analysis: A survey. **Pattern Recognition**, November 2002.
- FAUSETT, L. **Fundamentals of neural networks: architectures, algorithms, and applications**. [S.l.]: Prentice-Hall, Inc., 1994.
- GARBAS, J.-U. *et al.* Towards robust real-time valence recognition from facial expressions for market research applications. In: IEEE. **Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on**. [S.l.], 2013. p. 570–575.
- GARG, A.; BAJAJ, R. Facial expression recognition & classification using hybridization of ica, ga, and neural network for human-computer interaction. **Journal of Network**

Communications and Emerging Technologies (JNCET) www.jncet.org, v. 2, n. 1, 2015.

GONCALVES, A. R. **Máquina de Vetores de Suporte**. 2014. <http://www-users.cs.umn.edu/~andre/arquivos/pdfs/svm.pdf>, acessado em dezembro de 2016.

GRIGORESCU, S. E.; PETKOV, N.; KRUIZINGA, P. Comparison of texture features based on gabor filters. **IEEE Transactions on Image processing**, IEEE, v. 11, n. 10, p. 1160–1167, 2002.

GUHA, T. *et al.* On quantifying facial expression-related atypicality of children with autism spectrum disorder. In: IEEE. **2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)**. [S.l.], 2015. p. 803–807.

HAZELHOFF, L. *et al.* Behavioral state detection of newborns bases on facial expression analysis. **Advanced Concepts for Intelligent Vision systems**, 2009.

HECHT-NIELSEN, R. Theory of the backpropagation neural network. In: IEEE. **Neural Networks, 1989. IJCNN., International Joint Conference on**. [S.l.], 1989. p. 593–605.

HJELMÅS, E.; LOW, B. K. Face detection: A survey. **Computer vision and image understanding**, Elsevier, v. 83, n. 3, p. 236–274, 2001.

KATOH, A.; FUKUI, Y. Classification of facial expressions using self-organizing maps. **Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE**, v. 2, p. 986–989, 1998.

KOVÁCS, Z. L. **Redes neurais artificiais**. [S.l.]: Editora Livraria da Física, 2002.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2012. p. 1097–1105.

KUMAR, N. R.; BINDU, A. An efficient skin illumination compensation model for efficient face detection. In: IEEE. **IEEE Industrial Electronics, IECON 2006-32nd Annual Conference on**. [S.l.], 2006. p. 3444–3449.

LEE, B.; CHUN, J.; PARK, P. Classification of facial expression using svm for emotion care service system. In: IEEE. **Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2008. SNPD'08. Ninth ACIS International Conference on**. [S.l.], 2008. p. 8–12.

LEE, D.-T.; SCHACHTER, B. J. Two algorithms for constructing a delaunay triangulation. **International Journal of Computer & Information Sciences**, Springer, v. 9, n. 3, p. 219–242, 1980.

LEVI, G.; HASSNER, T. Emotion recognition in the wild via convolutional neural networks and mapped binary patterns. In: ACM. **Proceedings of the 2015 ACM on International Conference on Multimodal Interaction**. [S.l.], 2015. p. 503–510.

LIEN, J. J. *et al.* Automated facial expression recognition based on faces action units. **Third IEEE International Conference on Automatic Face and Gesture Recognition**, p. 390–395, 1998.

LITTLEWORT, G. *et al.* Towards social robots: Automatic evaluation of human-robot interaction by face detection and expression classification. **Advances in Neural Information Processing Systems**, 2003.

LITTLEWORT, G. C.; BARTLETT, M. S.; LEE, K. Automatic coding of facial expressions displayed during posed and genuine pain. **Image and Vision Computing**, Elsevier, v. 27, n. 12, p. 1797–1803, 2009.

LYONS, M. J.; BUDYNEK, J.; AKAMATSU, S. Automatic classification of single facial images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 21, n. 12, December 1999.

MICHIE, D.; SPIEGELHALTER, D. J.; TAYLOR, C. C. Machine learning, neural and statistical classification. Citeseer, 1994.

MOSES, Y.; REYNARD, D.; BLAKE, A. Determining facial expressions in real time. **Fifth International Conference on Computer Vision**, p. 296–301, 1995.

PADGETT, C.; COTTRELL, G. W. Representing face images for emotion classification. **Advances in neural information processing systems**, MORGAN KAUFMANN PUBLISHERS, p. 894–900, 1997.

PALI, V.; GOSWAMI, S.; BHAIYA, L. P. An extensive survey on feature extraction techniques for facial image processing. In: IEEE. **Computational Intelligence and Communication Networks (CICN), 2014 International Conference on**. [S.l.], 2014. p. 142–148.

PANTIC, M.; ROTHKRANTZ, L. J. Automatic analysis of facial expressions: The state of the art. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 12, December 2000.

PANTIC, M.; ROTHKRANTZ, L. J. Self-adaptive expert system for facial expression analysis. **IEEE International Conference on Systems, Man & Cybernetics**, v. 1, October 2000.

PARK, S.; KIM, D. Spontaneous facial expression classification with facial motion vectors. **8th International Conference on Automatic Face and Gesture Recognition**, September 2008.

PILOWSKY, I.; KATSIKITIS, M. The classification of facial emotions: a computer-based taxonomic approach. **Journal of Affective Disorders**, Elsevier, v. 30, n. 1, p. 61–71, 1994.

QUINLAN, J. R. **C4. 5: programs for machine learning**. [S.l.]: Elsevier, 2014.

SAATCI, Y.; TOWN, C. Cascaded classification of gender and facial expression using active appearance models. **7th International Conference on Automatic Face and Gesture Recognition**, April 2006.

SAMAL, A.; IYLANGAR, P. A. Automatic recognition and analysis of human faces and facial expressions: A survey. **Pattern Recognition**, v. 25, n. 1, June 1992.

SCHIAVENATO, M. *et al.* Neonatal pain expression: Evaluating the primal face of pain. **Pain**, v. 138, p. 240–271, 2008.

SU, L.; SATO, Y. Early facial expression recognition using early rankboost. **Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on**, p. 1–7, 2013.

SUWA, M.; SUGIE, N.; FUJIMORA, K. A preliminary note on pattern recognition of human emotional expression. **International joint conference on pattern recognition**, v. 1978, p. 408–410, 1978.

TERZOPOULOS, D.; WATERS, K. Analysis of facial images using physical and anatomical models. **Proceedings of the Third International Conference on Computer Vision**, v. 1, December 1990.

TIAN, Y.-L.; KANADE, T.; COHN, J. F. Facial expression analysis. In: **Handbook of face recognition**. [S.l.]: Springer, 2005. p. 247–275.

TURK, M.; PENTLAND, A. Eigenfaces for recognition. **Journal of cognitive neuroscience**, MIT Press, v. 3, n. 1, p. 71–86, 1991.

ULUKAYA, S.; ERDEM, Ç. E. A hybrid facial expression recognition method based on neutral face shape estimation. In: IEEE. **Signal Processing and Communications Applications Conference (SIU), 2012 20th**. [S.l.], 2012. p. 1–4.

VIOLA, P.; JONES, M. Robust real-time face detection. **International journal of computer vision**, Springer, v. 57, n. 2, p. 137–154, 2004.

WANG, S. *et al.* A natural visible and infrared facial expression database for expression recognition and emotion inference. **IEEE Transactions on Multimedia**, IEEE, v. 12, n. 7, p. 682–691, 2010.

WHITEHILL, J. *et al.* The faces of engagement: Automatic recognition of student engagement from facial expressions. **Affective Computing, IEEE Transactions on**, IEEE, v. 5, n. 1, p. 86–98, 2014.

WOLD, S.; ESBENSEN, K.; GELADI, P. Principal component analysis. **Chemometrics and intelligent laboratory systems**, Elsevier, v. 2, n. 1-3, p. 37–52, 1987.

YANG, M.-H.; KRIEGMAN, D. J.; AHUJA, N. Detecting faces in images: A survey. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, IEEE, v. 24, n. 1, p. 34–58, 2002.

YEASIN, M.; BULLOT, B.; SHARMA, R. Recognition of facial expressions and measurement of levels of interest from video. **Multimedia, IEEE Transactions on**, IEEE, v. 8, n. 3, p. 500–508, 2006.

ZHANG, X. *et al.* A high-resolution spontaneous 3d dynamic facial expression database. In: IEEE. **Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on**. [S.l.], 2013. p. 1–6.

ZHANG, X. *et al.* Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. **Image and Vision Computing**, Elsevier, v. 32, n. 10, p. 692–706, 2014.

APÊNDICE A – EXEMPLOS DE EXPRESSÕES FACIAIS SIMILARES

FIGURA 20: Tristeza classificada como Raiva

FONTE: A autora (2016)



FIGURA 21: Medo classificado como Felicidade

FONTE: A autora (2016)



FIGURA 22: Raiva classificada como Tristeza

FONTE: A autora (2016)