

Application for Emotion Recognition Through Facial Expressions for the Visually Disability

Adilmar Coelho Dantas
Faculty of Computer Science
Federal University of Uberlândia, UFU
Uberlândia, Minas Gerais, Brazil
akanehar@gmail.com

Alessandro Santana Martins
Campus Ituiutaba
Federal Institute of Triangulo Mineiro, IFTM
Ituiutaba, Minas Gerais, Brazil
alessandro@iftm.edu.br

Leandro Alves Neves
Department of Computer Science and Statistics
São Paulo State University, UNESP
São José do Rio Preto, São Paulo, Brazil
neves.leandro@gmail.com

Marcelo Zanchetta do Nascimento
Faculty of Computer Science
Federal University of Uberlândia, UFU
Uberlândia, Minas Gerais, Brazil
marcelo.zanchetta@gmail.com

Resumo—Emotions are sensations that stimulated can be expressed by representations of the face of a person. The absence of visual recognition of emotions can influence the social interactions of people with visual impairment. This work presents a mobile application capable of recognizing and classifying basic emotions (joy, sadness, anger, fear, disgust and surprise) in real time through facial expressions. These emotions are encoded in signs of sensory vibration so that it is possible to be used by people with visual impairment. This approach can identify and differentiate these emotions for social relationships. The developed interface was evaluated by obtaining a recognition rate of 88% for the six emotions. The experiments to evaluate two techniques for classifying these emotions in real time and the usability and suitability of visually impaired individuals.

Index Terms—Image Processing, Visual disability, Facial Expressions, Interfaces, Accessibility

I. INTRODUÇÃO

As emoções são sensações que quando provocadas por algum estímulo podem ser expressas por representações da face de um ser humano. A análise da face de um indivíduo pode ser considerada importante nas relações interpessoais, conforme descrito no trabalho de [2]. Segundo relatos de Ekman e Friesen (2003) [3], as expressões faciais de emoções básicas definidas por alegria, tristeza, raiva, medo, desgosto e surpresa podem ser inferidas por meio de músculos da face. Essas emoções são caracterizadas por mudanças na aparência da face que duram segundos ou frações de segundo para representação. Essas informações são visíveis e não invasivas na interpretação das emoções [3].

A comunicação não verbal, principalmente, na análise das emoções em deficientes visuais, apresenta ainda poucas pesquisas nos trabalhos presentes na Literatura [2]. Apesar das restrições visuais, normalmente, indivíduos com deficiência visual são capazes de evidenciar as expressões faciais das emoções. Isso mostra que esse indivíduo pode exibir as emoções na face embora ainda não tenha nenhum conhecimento visual das mesmas [1].

A ausência do reconhecimento visual das emoções pode provocar uma redução nas interações sociais. Isso ocorre principalmente pela dificuldade de inferir nas expressões faciais as emoções. Ferramentas computacionais que contribuam com o aperfeiçoamento do conhecimento não visual podem colaborar para a construção e aperfeiçoamento por meio de estímulos não visuais, como por exemplo, o som e o tato [17].

O presente trabalho mostra um *software*, denominado *Emotion Visual*, para auxiliar no processo de reconhecimento de emoções para os deficientes visuais. Para tanto, inicialmente, a detecção da face foi realizada com o algoritmo *Haar Cascade* [4]. Para realizar a localização e mapeamento dessas regiões foram utilizados 68 pontos projetados na face. Nessa etapa, foi utilizado a técnica de fluxo óptico para mapeamento desses pontos. Então, na classificação das emoções foi utilizada uma árvore de decisão. Para avaliação, foi utilizado o banco de imagens Cohn-Kanade (ck+), o qual tem imagens de cada uma das seis emoções básicas [10].

II. TRABALHOS CORRELATOS

Na Literatura existem diversos trabalhos de visão computacional que estudam a extração de características da face para a detecção e classificação das emoções por meio das expressões faciais. O trabalho de [12], propõe um sistema para detecção das sete emoções básicas (Neutro, Alegria, Tristeza, Raiva, Desgosto e Medo) em tempo real, utilizando o modelo *Piecewise Volume Deformation* (PVBD), baseado na teoria FACS. Os autores propõe, também a criação de um conjunto denominado Unidades de Movimento (UM) para uma melhor interpretação das Unidades de Ação (UA) no contexto computacional. Essa abordagem permite facilitar a interpretação e localização das regiões de interesse. Para a classificação foram utilizadas técnicas de aprendizagem de máquina, como *Naive Bayes*, e uma variação denominada TAN. Os resultados com o *Naive Bayes* foram considerados relevantes, se comparados com outros trabalhos da literatura.

Já os resultados apresentados para o algoritmo TAN não foram relevantes quando comparados com outros trabalhos, conforme mencionado pelos próprios autores.

Em [13] foi apresentado um sistema que permite a detecção de uma gama muito maior do comportamento facial, a partir dos movimentos dos músculos faciais das UA. Além disso, o trabalho apresenta informações de comportamentos temporais, ou seja, uma sequência de segmentos que contribuíram para a detecção da emoção. Para a localização correta das UA, o filtro Gabor foi empregado, permitindo a construção de uma máscara facial. Para a classificação, foi utilizada uma máquina de vetores de suporte (do inglês, *support vector machine* (SVM)). Os testes com o reconhecimento das UA em base de dados foram empregados e os resultados foram satisfatórios. Os resultados para a classificação das emoções, apesar de inferior aos resultados apresentados com as UA foram considerados importantes para área. Os testes realizados levaram em consideração emoções espontâneas e não espontâneas, outro aspecto de grande relevância em investigações para essa característica de sistema.

No trabalho de [14], um sistema foi apresentado baseado em uma abordagem para a classificação das expressões faciais usando modelos de aparência ativa (*Active Appearance Models*) com algoritmo (AAM) para as bordas da imagem. A aplicação dessa técnica resultou em mais robustez contra ruídos de iluminação variadas. Os dados obtidos permitiram uma melhor detecção das expressões faciais em condições desfavoráveis. Para a classificação, foram utilizados três técnicas diferentes, AAM, MLP e SVM. A aplicação foi avaliada usando a base de imagens FEEDTUM onde os melhores resultados obtidos foram o SVM com acurácia de 92,0% e MLP com valores de 75,0%. Os resultados demonstraram que utilização do classificador AAM não é suficiente para detectar as emoções de maneira satisfatória.

III. ARQUITETURA DA APLICAÇÃO

De maneira geral, o sistema proposto tem uma interface para detectar a face, processar as expressões faciais, classificar a emoção inferida e por último, fornecer um *feedback* para o deficiente visual. Neste caso, trata-se de uma resposta sensorial através da vibração. A Figura 1 traz o diagrama geral da interface desenvolvida e suas respectivas etapas de execução.

A. Detecção da Face

A primeira etapa a ser executada foi a detecção da face. Na face foram extraídas as características das regiões de interesse, tais como, olhos, boca, nariz e contorno. Para a detecção da face foi utilizado o algoritmo *Haar Cascade* [4], que através do cálculo da integral de imagem possui os melhores tempos de resposta para essas aplicações. Essa escolha ocorreu pelo fato desse algoritmo ser, computacionalmente, otimizado pelo classificador denominado Adaboost. Esse classificador otimiza o tempo de respostas e economiza recursos computacionais [5]. Durante a realização dos testes optou-se pela conversão da imagem no padrão de cores RGB para níveis de cinza. Isso

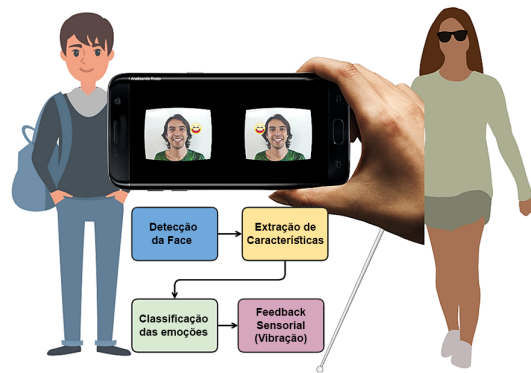


Figura 1. Diagrama representativo da interface e suas etapas de execução na detecção das emoções.

ocorreu com objetivo de reduzir o número de canais de cores a serem analisados nas imagens.

B. Extração de Características

Após a detecção da face, deve-se localizar as expressões faciais através das Unidades de Ação (UA) propostas por Ekman [3]. Para realizar a localização e mapeamento dessas regiões, foram utilizados 68 pontos projetados na face. Na Figura 2 é apresentado um exemplo do processo de captura das UA. Dentre os pontos projetados, um conjunto de 12 UA foram extraída para identificar as emoções inferidas através das expressões faciais. As UA empregadas neste estudo estão detalhadas na Tabela I.

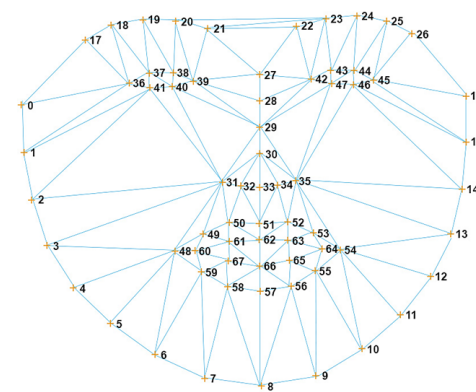


Figura 2. Máscara facial contendo os 68 pontos na face capturados pelas UA.

Para o mapeamento desses pontos foi utilizada a técnica de fluxo óptico. Essa técnica define que a intensidade entre quadros diferentes em uma sequência de imagens é aproximadamente constante em um intervalo de tempo pequeno, ou seja, o deslocamento será mínimo [6]. Após a determinação das UA, foi calculada a distância euclidiana entre os 12 pontos, conforme a Equação 1 para os pontos P e Q localizados na face. Essa informação foi repassada como parâmetro de entrada para a etapa responsável pela classificação das emoções.

Tabela I
UNIDADES DE AÇÃO EXTRAÍDAS NA FACE POR MEIO DA INTERFACE
PROPOSTA NESTE SISTEMA.

UA	UA Contraída - Movimento	Pares Associados
1	vertical superior do centro do lábio	(P51-P66)
2	vertical inferior do centro do lábio	(P57-P66)
3	horizontal do canto esquerdo da boca	(P48-P66)
4	vertical do canto esquerdo da boca	(P48-P31)
5	horizontal do canto direito da boca	(P54-P64)
6	vertical do canto direito da boca	(P54-P35)
7	vertical da sobrancelha direita	(P24-P44)
8	vertical da sobrancelha esquerda	(P19-P37)
9	Enrugador de Nariz do canto direito	(P35-P53)
10	Enrugador de Nariz do canto esquerdo	(P31-P49)
11	Piscar olho direito	(P44-P46)
12	Piscar olho esquerdo	(P37-P41)

$$P = (p_1, p_2, \dots, p_n) \quad e \quad Q = (q_1, q_2, \dots, q_n)$$

$$\sqrt{(p_1 - q_1)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

C. Classificação das Emoções

Com as expressões faciais identificadas na face, ocorreu a classificação das emoções. Para a classificação das emoções foi utilizada uma árvore de decisão. A árvore de decisão é uma estrutura composta por três elementos básicos, sendo eles: nós de decisão, que correspondem aos atributos; bordas ou ramos, que correspondem aos diferentes valores de atributos possíveis e as folhas, que incluem objetos que pertencem à mesma classe ou que são muito semelhantes [7]. A escolha dessa técnica foi baseada no trabalho de Dantas et al. (2015) [8], devido à solução proporcionar uma boa compreensão visual e por possibilitar a reconstrução de maneira adequada as mudanças aos músculos da face analisada.

Para a construção da árvore de decisão foi utilizado o algoritmo ID3, um método que permite gerar árvores relativamente pequenas quando comparado com outras técnicas. Esse método também está entre os mais conhecidos e bem sucedidos algoritmos de inferência indutiva [16], o qual é aplicado em diversas áreas do conhecimento.

O algoritmo básico do ID3 constrói árvores de decisão partindo da raiz. Essa raiz é determinada avaliando cada atributo por um método estatístico para determinar como esse classifica os exemplos de treinamento. O melhor atributo é selecionado como nó raiz da árvore. Um descendente do nó raiz é criado para todos os possíveis valores desse atributo. Os exemplos de treinamento são particionados e associados a cada nó descendente com a finalidade de selecionar o melhor atributo. Essa característica gera uma procura de uma árvore de decisão aceitável no qual o método nunca retrocede às escolhas feitas anteriormente.

D. Aplicativo Móvel

Após a etapa de classificação das emoções, um algoritmo foi construído para gerar os sinais de vibração no aparelho do

deficiente visual. Nesse algoritmo, cada emoção corresponde a uma intensidade de vibração. Essa intensidade ajuda o deficiente visual a treinar e reconhecer cada uma das emoções, conforme exibido a Figura 3.



Figura 3. Interface para treinar o reconhecimento das emoções a partir da vibração.

A ferramenta foi desenvolvida por meio da plataforma Unity, uma ferramenta de desenvolvimento de jogos 2D e 3D, utilizada para a criação de jogos multiplataforma e que faz uso de realidade virtual e aumentada [9].

A interface pode funcionar em dois modos: *online* e *offline*, devido à implementação de um banco de dados baseado em computação em nuvem. Toda vez que o aplicativo é executado pela primeira vez, obtém as informações necessárias para a classificação das emoções, e as demais etapas relacionadas a processamento de imagem e extração de características são executados na unidade de processamento gráfico do equipamento.

Todos os dados processados são armazenados em estruturas de dados no próprio aparelho e quando identificado uma conexão com a internet, a aplicação realiza a sincronização junto ao banco de dados hospedado em um serviço na nuvem. Essa estratégia permite que essas informações sejam utilizadas para aperfeiçoar a taxa de classificação da aplicação para as emoções básicas através da reconstrução do modelo computacional.

IV. RESULTADOS

Para avaliar a aproximação proposta foram realizados experimentos com o objetivo de avaliar a taxa de acurácia no reconhecimento das emoções através das expressões faciais. Nesse experimento foi investigado a aproximação proposta e o método apresentado em [8]. A técnica apresentada em [8] é uma solução proposta pela Universidade de Oxford juntamente com a empresa Microsoft, para a classificação das emoções [15].

Nesta etapa, foi utilizado o banco de imagens Cohn-Kanade (ck+), com imagens de cada uma das seis emoções básicas [10]. Essa base possui também imagens de diferentes etnias e sexo. Além disso, essas imagens foram organizadas em uma sequência que representam cada emoção no seu estado do mínimo ao máximo. A base possui um total de 1452 imagens em escala de cinza separadas por emoção, conforme a Tabela II.

Tabela II
DISTRIBUIÇÃO DAS IMAGENS POR EMOÇÃO NA BASE COHN-KANADE.

Emoção	Quantidade de Imagens
Alegria	324
Tristeza	253
Raiva	183
Medo	182
Desgosto	182
Surpresa	328

Para o treinamento dos classificadores, foram selecionadas dez imagens de maneira aleatória sem repetição para cada uma das seis emoções básicas. As duas soluções exploradas foram submetidas aos testes para reconhecer a face e o desempenho das abordagens. Na Figura 4 são apresentados os resultados para cada uma das seis emoções da técnica proposta baseado em [8]

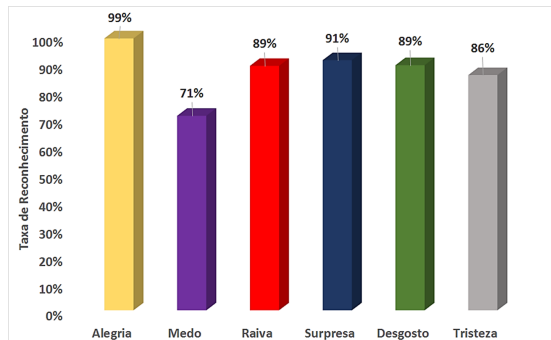


Figura 4. Taxa de acerto para cada uma das seis emoções, através da técnica proposta por [8].

Analisando os resultados apresentados na Figura 4, foi possível obter uma taxa média de acertos de 88% para a classificação das emoções. Esse valor de reconhecimento pode ser considerado relevante, pois detecta e classifica as emoções em tempo real a partir das expressões faciais. Isso é expressivo uma vez que muitos estudos presentes na Literatura definem que essa tarefa é algo complexo devido ao grande número de detalhes, características e interferências que podem ocorrer durante o processo de detecção das informações [11]. Além disso, nota-se que em apenas uma das emoções básicas, o medo, os resultados não foram expressivos. Ressalta-se que esse problema ocorre devido ao fato de que a emoção medo possui muitas UA em comum com outras emoções, o que dificulta a classificação por meio do algoritmo, conforme relatado no trabalho de [8].

Na Figura 5 são apresentados os resultados obtidos nos testes realizados utilizando a API da Microsoft.

Nos testes realizados com a API da Microsoft, os resultados também foram satisfatórios, mas com valores de acurácia para as emoções medo (5%) e desgosto (41%) foram considerados baixos em relação às outras emoções. Nota-se que esses valores foram discrepantes em relação ao método de [8], o qual a emoção medo foi de 71% e desgosto 89% de acurácia. Essa limitação na classificação dessas emoções podem estar

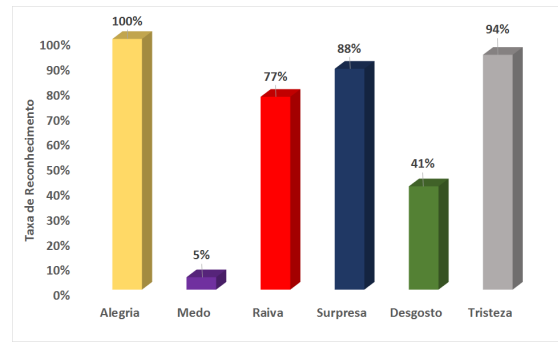


Figura 5. Resultados da taxa de acerto para a API da Microsoft.

relacionados às características da API, pois essa solução não considera informações das UA na extração de características.

Para avaliar a usabilidade e as contribuições da interface desenvolvida, foi elaborado um questionário com quatro perguntas para indivíduos. Esse questionário foi avaliado por um grupo com um participante com deficiência visual e por nove participantes sem a deficiência, os quais tiveram sua visão vendada durante a avaliação do método. Após a utilização da ferramenta, as questões foram empregadas com o objetivo de obter o *feedback* dos usuários para a realização de possíveis adaptações ou alterações no aplicativo. As perguntas elaboradas para o questionário estão disponíveis na Tabela III. Esse trabalho foi submetido ao comitê de ética da Universidade Federal de Uberlândia para aprovação e aplicação dentro das normas vigentes para as pesquisas que envolvem seres humanos.

Tabela III
QUESTIONÁRIO PARA A AVALIAÇÃO DA INTERFACE.

Questões	
Foi fácil entender como utilizar a interface?	-2() -1() 0() 1() 2()
Você considera importante as emoções em nossas vidas?	-2() -1() 0() 1() 2()
Você conseguiu identificar bem as emoções?	-2() -1() 0() 1() 2()
Você usaria a interface em alguma situação?	-2() -1() 0() 1() 2()

O gráfico presente na Figura 6 traz as respectivas respostas dos participantes para cada um dos quatro questionamentos. Essa avaliação permitiu mostrar que os participantes reconhecem a importância das emoções nas relações intersociais e consideraram relativamente simples a utilização da interface para o reconhecimento das emoções em tempo real, mesmo com as limitações da visão.

V. CONCLUSÕES

Este trabalho apresentou e avaliou um software para a detecção e classificação das emoções através das expressões faciais em dispositivos móveis. Além disso, o método proposto foi empregado às pessoas com deficiência visual. Os experimentos realizados permitiram avaliar duas propostas para a classificação das emoções através da base imagens Cohn-Kanade [10] e um grupo de voluntários. Os resultados

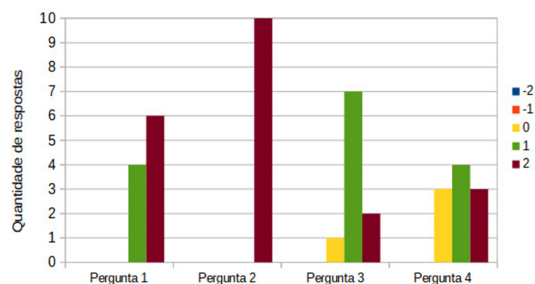


Figura 6. Gráfico da taxa de acerto para cada uma das seis emoções.

foram satisfatórios e relevantes para a implementação em aplicações de reconhecimento de emoções. O desenvolvimento e aplicação desse software permitiu também observar que técnicas de computação afetiva e processamento digital de imagens podem ser empregadas com o objetivo de fornecer melhor qualidade de vida para portadores de deficiência visual. Isso possibilita aos indivíduos com limitações visuais a possibilidade de reconhecer as seis emoções básicas através da interface. A solução desenvolvida ainda apresenta algumas dificuldades para a localização e classificação das emoções a partir das expressões faciais na presença de oclusões em câmeras de baixa resolução e mudanças em relação a emoção avaliada.

Em trabalhos futuros, pretende-se investigar novas técnicas de processamento de imagem que possam melhorar as imagens numa etapa de pre-processamento. Ainda, será realizada a avaliação das técnicas exploradas em outras bases de imagens e a comparação com outras soluções existentes.

VI. AGRADECIMENTOS

Os autores gostariam de agradecer ao CNPQ (Projeto 427114/2016-0) pelo suporte financeiro.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 1744339.

REFERÊNCIAS

- [1] CASTANHO, A. R. S. P. et al. Gestos não-verbais espontâneos e canais de expressão emocional voluntária em deficientes visuais. *Revista Benjamin Constant*, v. 26, 2003.
- [2] FERREIRA, Bárbara Carvalho et al. Expressões faciais de emoções de crianças com deficiência visual e videntes: avaliação e intervenção sob a perspectiva das Habilidades Sociais. 2012. Tese em Psicologia – Universidade Federal de São Carlos - UFSCar.
- [3] EKMAN, Paul; FRIESEN, Wallace V. *Unmasking the face: A guide to recognizing emotions from facial clues*. Ishk, 2003.
- [4] Viola, P., Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on (Vol. 1, pp. I-1)*. IEEE.
- [5] FREUND, Yoav; SCHAPIRE, Robert; ABE, Naoki. A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, v. 14, n. 771-780, p. 1612, 1999.
- [6] HORN, Berthold KP; SCHUNCK, Brian G. Determining optical flow. *Artificial intelligence*, v. 17, n. 1-3, p. 185-203, 1981.
- [7] RUSSELL, Stuart J.; NORVIG, Peter. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited., 2016.

- [8] DANTAS, Adilmar Coelho et al. Reconhecimento dinâmico de emoções através de expressões faciais utilizando árvore de decisão. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. 2015. p. 1102.
- [9] Unity. *User Manual*. 2018. Disponível em: <https://docs.unity3d.com/Manual/index.html>; Acesso em: 01 abr. 2018.
- [10] LUCEY, Patrick et al. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010. p. 94-101.
- [11] KONAR, Amit; CHAKRABORTY, Aruna. *Emotion recognition: A pattern analysis approach*. John Wiley Sons, 2014.
- [12] AZCARATE, Aitor et al. *Automatic facial emotion recognition*. Universiteit van Amsterdam, p. 1-6, 2005.
- [13] VALSTAR, Michel F.; PANTIC, Maja. Fully automatic recognition of the temporal phases of facial actions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, v. 42, n. 1, p. 28-43, 2012.
- [14] MARTIN, Christian; WERNER, Uwe; GROSS, Horst-Michael. A real-time facial expression recognition system based on active appearance models using gray images and edge images. In: *Automatic Face Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE, 2008. p. 1-6.
- [15] Microsoft, "Cognitive Services" <https://azure.microsoft.com/pt-br/services/cognitive-services/emotion/>, 2018 (Acessado em 01 de Junho de 2018).
- [16] NORVIG, Peter; RUSSELL, Stuart. *Inteligência Artificial: Tradução da 3a Edição*. Elsevier Brasil, 2014.
- [17] MATSUMOTO, David; WILLINGHAM, Bob. Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *Journal of personality and social psychology*, v. 96, n. 1, p. 1, 2009.