

INTERFACE MÓVEL PARA O RECONHECIMENTO DE EMOÇÕES ATRAVÉS DE EXPRESSÕES FACIAIS EM DEFICIENTES VISUAIS

A. C. Dantas*, E. B. Gouveia**, M. C. Melo**, C. A. Gonçalves* e M. Z. Nascimento*

*Universidade Federal de Uberlândia – Faculdade de Computação, Uberlândia, Brasil

** Universidade Federal de Uberlândia – Faculdade de Engenharia Elétrica, Uberlândia, Brasil

e-mail: akanehar@gmail.com

Resumo: Este trabalho apresenta uma interface móvel capaz de reconhecer e classificar as emoções básicas (alegria, tristeza, raiva, medo, desgosto e surpresa) em tempo real, através das expressões faciais. Essas emoções são codificadas em sinais sensíveis de vibração para que seja possível a utilização por indivíduos com deficiência visual, proporcionando assim a capacidade de identificar e diferenciar essas emoções em suas relações intersociais. A interface desenvolvida foi avaliada obtendo uma taxa de reconhecimento de 88% para as seis emoções, além disso foram realizados experimentos para avaliar a usabilidade e adequação com indivíduos com deficiência visual.

Palavras-chave: Emoções, Deficiência Visual, Expressões Faciais, Interfaces, Acessibilidade, Processamento de Imagens.

Abstract: *Abstract: On this paper is presented a mobile interface 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 sensory vibration signals so that it is possible for use by visually impaired individuals, thus providing the ability to identify and differentiate these emotions in their social relationships. The developed interface was evaluated obtaining a recognition rate of 88% for the six emotions, in addition experiments were performed to evaluate the usability and adequacy with individuals with visual impairment.*

Keywords: *Emotions, Visual disability, Facial Expressions, Interfaces, Accessibility, Image Processing.*

Introdução

A comunicação não verbal, principalmente, no que diz respeito ao conhecimento das emoções em deficientes visuais, apresenta poucas pesquisas na literatura apesar da sua grande importância nas relações sociais do cotidiano [2]. Apesar das limitações visuais, o indivíduo é capaz de evidenciar as expressões faciais das emoções básicas. Isso mostra que o deficiente visual congênito exibe emoções na face embora não tenha nenhum conhecimento visual [1].

As emoções são expressas através das informações faciais, mas não somente ou exclusivamente. Além disso, a análise da face é considerada importante para os processos sociais [2], pois é uma das formas de interações, inclusive para interpretar as emoções do interlocutor. Segundo Ekman e Friesen (2003) as expressões faciais de

emoções básicas (alegria, tristeza, raiva, medo, desgosto, surpresa) podem ser inferidas através de sinais rápidos da face, que são caracterizados por mudanças na aparência da face que duram segundos ou frações de segundo. Esses sinais são mais visíveis que outros existentes e menos invasivos para interpretar essas emoções de maneira natural [3].

As emoções são importantes para as interações sociais e os deficientes visuais tem limitações devido sua acuidade visual. Ferramentas com a proposta de ajudar essas pessoas são poucas para esse grupo de indivíduos. O presente trabalho apresenta o *software* denominado Emotion Visual, o qual emprega técnicas de processamento de imagens e inteligência artificial em uma interface móvel desenvolvida para tornar o reconhecimento das emoções através das expressões faciais uma realidade.

Materiais e métodos

Para o desenvolvimento desta interface móvel capaz de realizar a intermediação entre as expressões faciais inferidas pelo interlocutor durante um diálogo comum e o deficiente visual, diversas tecnologias foram empregadas. De maneira geral uma interface com esse objetivo deve ser capaz primeiramente de detectar a face, processar as expressões faciais presentes na face, classificar corretamente a emoção inferida no presente momento 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.

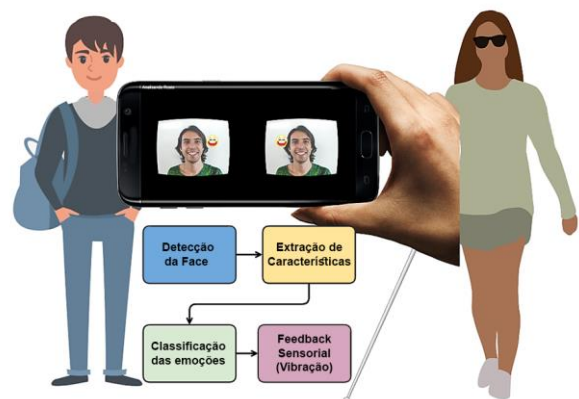


Figura 1: Diagrama representativo da interface e suas etapas de execução.

UA	Descrição
1	Movimento vertical superior do centro do lábio
2	Movimento vertical inferior do centro do lábio
3	Movimento horizontal do canto esquerdo da boca
4	Movimento vertical do canto esquerdo da boca
5	Movimento horizontal do canto direito da boca
6	Movimento vertical do canto direito da boca
7	Movimento vertical da sobrancelha direita
8	Movimento vertical da sobrancelha esquerda
9	Enrugador de Nariz do canto direito
10	Enrugador de Nariz do canto esquerdo
11	Piscar olho direito
12	Piscar olho esquerdo

A primeira etapa a ser executada é a correta detecção da face. Além disso é na face são extraídas as características das regiões de interesse (olhos, boca, nariz e contorno). Para a correta 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 também ocorreu pelo fato deste algoritmo ser computacionalmente otimizado pelo classificador denominado *Adaboost*. Esse classificador otimiza o tempo de respostas [5] e economiza recursos computacionais. Durante a realização dos testes optou-se pela conversão da imagem em níveis de cinza, para reduzir a quantidade de ruídos presentes nas imagens.

Após a correta detecção da face, então, deve-se localizar corretamente as expressões faciais através das Unidades de Ação (UA) propostas por Ekman [3]. Para realizar a localização e mapeamento dessas regiões foi utilizado 68 pontos projetados na face, conforme imagem na Figura 2. Dentre os pontos projetados um conjunto de 12 UA foi extraída para que fosse possível identificar as emoções inferidas através das expressões faciais. As UA, empregadas neste estudo estão detalhadas na Tabela 1.

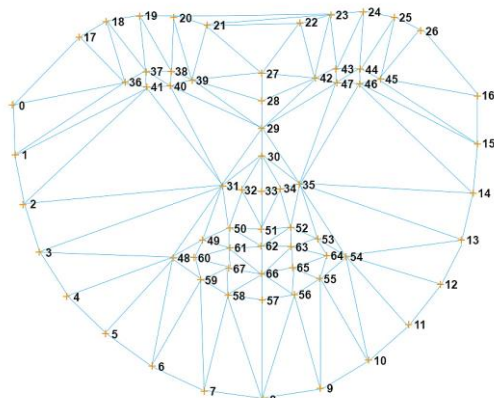


Figura 2: Máscara facial contendo os 68 pontos na face.

Tabela 1: Unidades de Ação extraídas na face através da interface.

Para o correto mapeamento desses pontos foi utilizado 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, em um pequeno intervalo de tempo o deslocamento será mínimo [6]. No fim da determinação das UA's, foi calculado a distância euclidiana entre os 12 pontos, conforme a equação (1) para os pontos P e Q localizados na face. Essa informação foi então repassada como parâmetro de entrada para a etapa responsável pela classificação das emoções.

$$P = (p1, p1, \dots, pn) \quad Q = (q1, q2, \dots, qn) \\ \sqrt{\sum_{i=1}^n (pi - qi)^2} \quad (1)$$

Com as expressões faciais identificadas na face, ocorreu a classificação das emoções. Para a classificação das emoções foi utilizado 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 em (Dantas, 2015) [8], devido a solução proporcionar uma boa compreensão visual e por possibilitar a reconstrução de maneira adequada e adaptada constantemente pela mudança dos músculos da face analisada.

Assim após a classificação da emoção, um algoritmo foi construído para gerar os sinais de vibração no aparelho do deficiente visual. Nesse algoritmo cada emoção corresponde uma intensidade de vibração. Essa intensidade ajuda o deficiente visual treinar a 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 plataforma 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 a 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 os demais processamentos de imagem e extração de características são executados na *Graphics Processing Unit*, ou Unidade de Processamento Gráfico (GPU) do aparelho. 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.

Resultados

Para avaliar a interface desenvolvida foi realizado experimentos, o primeiro com o objetivo de avaliar a acurácia da aplicação desenvolvida para reconhecer as emoções através das expressões faciais, e o segundo para medir a usabilidade por meio de questionários aplicados com os usuários.

O experimento para investigar a capacidade do sistema em detectar as emoções foi realizado usando o banco de imagens Cohn-Kanade [10], com imagens de cada uma das seis emoções básicas. Essa base possui também imagens de diferentes etnias e sexo. Além disso, essas imagens estão em 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 2.

Tabela 2: Quantidade de imagem por emoção presente na base de imagens Cohn-Kanade.

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

A aplicação desenvolvida foi submetida a testes para reconhecer a face e classificar as respectivas emoções em cada uma das imagens, permitindo assim avaliar a acurácia do algoritmo desenvolvido. Na Figura 4 temos o gráfico com os respectivos resultados para cada uma das seis emoções.

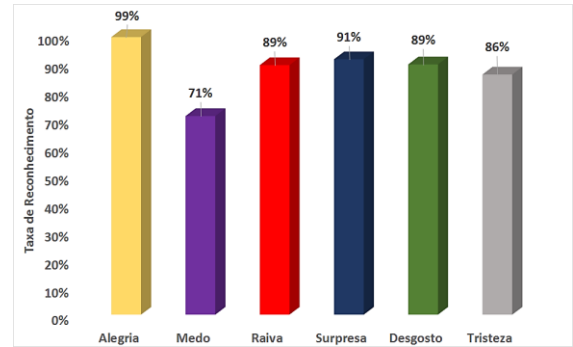


Figura 4: Gráfico da taxa de acerto para cada uma das seis emoções.

Analisando os resultados apresentados no gráfico foi possível obter uma taxa média de acertos de 88% para a classificação das emoções. Baseado em resultados de trabalhos na área, esse valor de reconhecimento pode ser considerado relevante, pois detecta e classifica as emoções em tempo real a partir das expressões faciais. Isso é relevante uma vez que muitos estudos presentes na literatura mostram 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].

Para avaliar a usabilidade, e as contribuições da interface desenvolvida foi elaborado um questionário com quatro perguntas. Esse questionário foi avaliado por um grupo com um participante com deficiência visual e por nove participantes sem a deficiência, mas vendados durante todo experimento. Após a utilização da ferramenta, as questões foram empregadas com o objetivo 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 3 abaixo.

Tabela 3: 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()

Essa avaliação permitiu concluir que os participantes reconhecem a importância das emoções nas relações intersociais e consideraram relativamente simples a utilização das interfaces para o reconhecimento das emoções em tempo real, mesmo com ausência da visão. O gráfico da Figura 5 traz as respectivas respostas dos participantes para cada um dos quatro questionamentos.

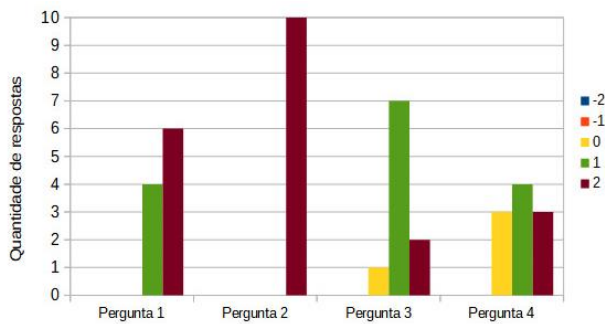


Figura 5: Respostas dos questionários de avaliação da interface.

Além da avaliação qualitativa realizada por meio dos questionários foi realizado também uma avaliação quantitativa, esta análise teve como objetivo avaliar a capacidade dos participantes em identificar três emoções básicas em três seções, através da utilização do aplicativo e por meio dos sinais sensitivos. Através dos experimentos realizados foi possível obter uma taxa média de acerto de 6,4.

Discussão

Os testes realizados para avaliar o desempenho das técnicas computacionais empregadas para a classificação das seis emoções básicas foram satisfatórios para a implementação na interface. Durante a realização dos experimentos foi observado também que a emoção medo obteve uma taxa inferior de classificação, devido ao fato desta emoção possuir uma grande quantidade de UA, sendo necessário uma quantidade maior de imagens para treinamento.

Conclusão

O desenvolvimento e aplicação deste trabalho permitiu 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, proporcionando a esses indivíduos a possibilidade de conhecer as seis emoções básicas através da interface desenvolvida e utilizar essa capacidade para criar e manter as relações intersociais. Como trabalho futuro pretende-se implementar novas emoções e sentimentos na interface e avaliar novas técnicas para a detecção e classificação das emoções através das expressões faciais.

Agradecimentos

Gostaríamos de agradecer ao Laboratório Interdisciplinar de Processamento e Análise de Imagens – LIPAI e a CAPES.

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.

[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-I). 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.