

Aplicação para reconhecimento dinâmico de emoções em ambientes virtuais de aprendizagem

Adilmar C. Dantas¹, Sara L. Melo¹, Márcia A. Fernandes¹, Eduardo K. Takahashi²

¹Faculdade de Computação – Universidade Federal de Uberlândia (UFU)

²Instituto de Física – Universidade Federal de Uberlândia (UFU)

{adilmar, sara}@mestrado.ufu.br, {marcia, ektakahashi}@ufu.br

Resumo. Este artigo descreve um sistema de reconhecimento dinâmico das sete emoções básicas (alegria, surpresa, raiva, medo, desgosto, tristeza e neutra) para ambientes de educação à distância. Para isto, a detecção das expressões faciais foi feita através da inserção de pontos na face tendo como base as unidades de movimento. Após a extração de características na face encontrada, foi utilizada a técnica *Árvore de Decisão* para a classificação precisa das emoções. A ferramenta desenvolvida foi integrada ao Moodle podendo ser acessada também por dispositivos móveis. O reconhecimento das emoções foi testado e validado com imagens estáticas e vídeos gerados da base de dados *Cohn-Kanade*. Os resultados obtidos foram satisfatórios para todas as emoções, atingido um percentual geral de 86,4% de acurácia.

1. Cenário de Uso

O aplicativo desenvolvido neste trabalho está inserido na plataforma Moodle do Núcleo de Pesquisa em Tecnologias Cognitivas (NuteC) da Universidade Federal de Uberlândia (UFU) e será utilizado em um contexto real de ensino-aprendizagem, na modalidade *e-learning*, ao longo do segundo semestre letivo de 2015, na disciplina Introdução ao Eletromagnetismo, que é um componente curricular obrigatório dos cursos de Física da UFU. A intenção é detectar as emoções em tempo real dos estudantes que estiverem acessando as atividades disponibilizadas na plataforma Moodle, sejam nas aulas presenciais ou não, com intuito de reconhecer o impacto que uma informação apresentada pelo professor produz ao estudante. O conhecimento da reação do aluno pode ser um fator auxiliar no processo de ensino-aprendizagem, uma vez que os alunos raramente manifestam explicitamente suas dúvidas, incertezas, desconfiças, medos, surpresa, etc. durante uma aula. O reconhecimento de alguma emoção negativa possibilitará uma intervenção do professor no sentido de buscar maior eficácia no processo de mediação entre o aluno e o saber a ser aprendido.

2. Desenvolvimento

Para o desenvolvimento desta aplicação capaz de detectar e classificar em tempo real as sete emoções básicas dos estudantes em ambientes de ensino à distância (EAD) foram desenvolvidos e validados os seguintes passos descritos abaixo.

2.1. Detecção da Face

Antes de qualquer processamento a ser realizado neste trabalho deve-se primeiramente realizar a detecção da face do estudante e das regiões de interesse (olhos, boca, nariz e

contorno da face), esta detecção foi feita através da utilização de classificadores desenvolvidos com auxílio do classificador em cascata proposto por [Viola and Jones 2004]. Este classificador contém parâmetros para localizar faces em imagens e vídeos de forma rápida, devido ao fato de ser otimizado computacionalmente pelo uso do algoritmo *Ada-boost* [Schapire 1990] em seu treinamento.

A detecção da face neste trabalho é tão importante quanto a detecção das unidades de ação, pois sem esta etapa não seria possível extrair informações das expressões faciais inferidas pela face do estudante. Por se tratar de uma etapa extremamente importante realizou-se o tratamento de oclusões na imagem como por exemplo: barba, bigode e óculos, onde tais oclusões poderiam interferir na detecção da face, para isso construiu-se classificadores em cascata para tratamento exclusivo dessas oclusões com a finalidade de minimizar estes possíveis ruídos.

2.2. Detecção das Unidades de Ação

Após a detecção da face o próximo passo desenvolvido foi o processamento das regiões de interesse, nestas regiões de interesse deve-se extrair como conhecimento as unidades de ação propostas por [Ekman and Friesen 1978]. Para localização e mapeamento dessas regiões utilizou-se pontos projetados na face, estes pontos são mapeados na face através da técnica de Fluxo Óptico [Barron et al. 1994] que é responsável por manter o acompanhamento cinemático das regiões de interesse.

Após esta detecção foram identificados e mapeados 12 pontos nas regiões de interesse da face adquirida, estes pontos foram inseridos conforme as unidades de movimento UM's propostas por [Azcarate et al. 2005] exibidos na Figura 1.

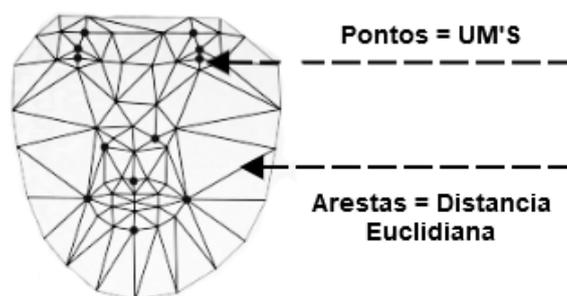


Figura 1. Máscara facial com os 12 pontos inseridos nas regiões de interesse

Os pontos foram inseridos na face conforme detalhados na Figura 1, sendo assim, ao detectar a face são inseridos 4 pontos na região da boca, 2 pontos na região do nariz, 4 pontos na região dos olhos e, 2 pontos na região da sobrancelha. Através destes pontos calculou-se a Distância Euclidiana entre os pontos referentes às UM's. Este cálculo foi utilizado para determinar se uma determinada UM que está relacionada à UA foi inferida em um dado instante.

2.3. Classificação das Emoções

Para a classificação das emoções foi utilizado a técnica Árvore de Decisão, pelo fato da árvore gerar uma melhor compreensão visual da sua estrutura facilitando assim uma melhor acurácia na classificação das emoções.

O processo de indução em Árvore de Decisão é uma das formas mais simples e, mais bem-sucedidas, de algoritmos de aprendizagem de máquina. Assim, tem-se como entrada um objeto ou situação descritos por um conjunto de atributos e retorna uma decisão, ou seja, o valor de saída previsto de acordo com a sua entrada.

Uma Árvore de Decisão alcança sua decisão executando uma sequência de testes. Cada nó interno na árvore corresponde a um teste do valor de uma das propriedades, e as ramificações a partir do nó são identificadas como os valores possíveis do teste. Cada nó de folha na árvore especifica o valor a ser retornado se aquela folha for alcançada. A representação de árvores de decisão parece ser muito natural para os seres humanos e utilizadas em diversas áreas do conhecimento [Russell and Norvig 2004].

Para a classificação das emoções através desta técnica computacional cada nó corresponde a determinada emoção que deseja classificar, os ramos são compostos pelo cálculo das distâncias euclidianas referente às emoções básicas e, por fim as folhas representam a emoção induzida à partir dos ramos da árvore.

2.4. Integração ao Ambiente Moodle

A plataforma Moodle é uma plataforma de aprendizagem projetada para fornecer aos educadores, administradores e estudantes um único sistema robusto, seguro e integrado para criar ambientes de aprendizagem personalizados [Moodle].

Para realizar a integração desta ferramenta de reconhecimento dinâmico de emoções ao Moodle foi desenvolvido um pequeno módulo (pacotes de configurações específicas) no qual o administrador do ambiente virtual é capaz de instalar o módulo responsável pela detecção das emoções. Este módulo consiste em pequenas alterações realizadas para permitir que os *scripts* de detecção e classificação de emoções possa comunicar, armazenar e compartilhar informação com a base de dados do ambiente Moodle, esta integração é exemplificada pelo diagrama na figura 2.

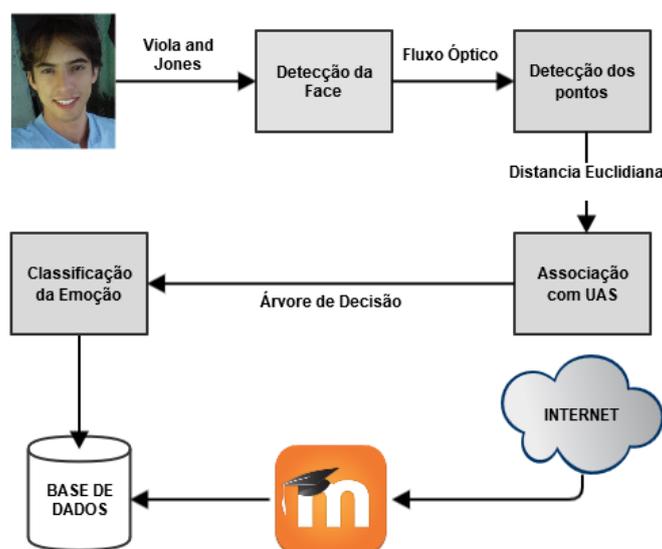


Figura 2. Diagrama: Módulo de detecção de emoções integrado ao ambiente Moodle

A segunda parte deste módulo é responsável pela geração de relatórios, estes relatórios podem ser personalizados de acordo com a necessidade do instrutor do curso, tais como: relatório emocional da turma ou relatório por aluno em forma de gráficos gerados dinamicamente a partir da base de dados em comum com o ambiente virtual de aprendizagem.

2.5. Testes e Validação do Sistema

Para o treinamento da Árvore de Decisão foi desenvolvido um conjunto de testes composto por 20 imagens extraídas da base de dados [Kanade and Cohn 2005] relacionadas a cada uma das emoções. Essa base possui imagens de diferentes etnias e sexo, além disso, as imagens estão em sequências que representam cada emoção no seu estado inicial até inferir o máximo de cada emoção.

A seleção das imagens para o treinamento foi feita utilizando um agrupamento das imagens do início, meio e fim da sequência que representa a emoção. Esta escolha foi utilizada para evitar que a árvore de decisão construa pesos somente para a emoção em seu estado máximo ou mínimo, que poderia levar a uma classificação errônea ou com baixa acurácia.

Para efetuar testes com imagens foram utilizadas 1452 imagens. Para os testes com vídeos foi necessário um trabalho minucioso na base de dados [Kanade and Cohn 2005]. Primeiramente, separou imagens por emoção e por pessoa e, em seguida foi construído uma sequência de *frames* que compõe cada vídeo. Assim, foram gerados 194 vídeos, a Tabela 1 exibe a quantidade de imagens e vídeos referente a cada emoção utilizadas para testes de classificação das emoções no modelo proposto.

Tabela 1. Total de imagens e vídeos gerados para cada emoção

Emoção	Imagens	Vídeos
Alegria	324	43
Tristeza	253	34
Raiva	183	22
Surpresa	328	43
Desgosto	182	26
Medo	182	26
Total	1452	194

Através das imagens e vídeos relacionadas a cada emoção o sistema foi implementado e, a validação foi feita a partir de testes feitos com imagens e vídeos. Os resultados obtidos para classificação das emoções para os dois testes são exibidos na seção resultados da validação .

3. Apresentação do Software

A aplicação desenvolvida foi integrada ao ambiente Moodle e é possível acessar a aplicação experimental através do seguinte endereço <http://inteligenciaartificial.facom.ufu.br/face-uas/> sob licença de uso livre, o módulo responsável pela integração no Moodle estará disponível através do mesmo endereço

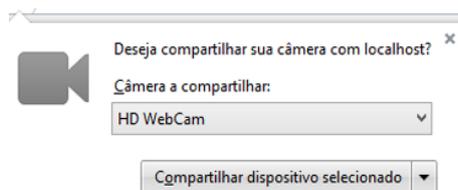


Figura 3. Interface inicial da aplicação, solicitação Webcam.

para *download* de forma gratuita. Ao acessar o endereço acima a seguinte interface será exibida conforme na figura 3.

Nesta primeira interface o usuário deve permitir o acesso a *webcam* que é responsável pela transmissão das imagens para efetuar a classificação das emoções através da expressão facial. Após esta permissão a aplicação será responsável pela detecção da face e a plotagem dos pontos para em seguida realizar os cálculos necessários para determinar a emoção inferida conforme demonstrado na figura 4.

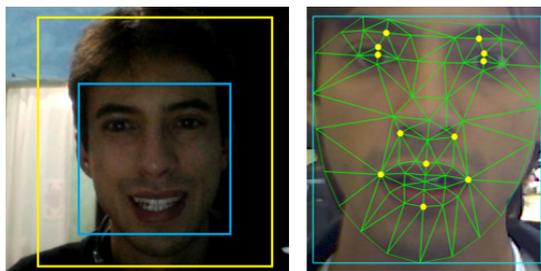


Figura 4. Detecção da face e plotagem dos pontos nas respectivas UM's.

A região localizada a esquerda do usuário conforme a figura 5 é apresentada as emoções inferidas pelo estudante, além disso é detectado o nível de atenção do usuário ao conteúdo proposto, caso este nível de atenção seja considerado inferior a aplicação emite um alerta para o usuário informando sua falta de atenção durante a atividade proposta. A sua direita é exibido um gráfico que é atualizado em tempo real com cores distintas para exibir os níveis de cada uma das seis emoções (alegria, raiva, surpresa, desgosto, medo, tristeza) inferidas temporalmente.



Figura 5. Aplicação em execução web

Além disso, é possível acessar a aplicação por meio de computadores e dispositivos móveis devido as tecnologias *Web* utilizadas na construção da aplicação que concentra todos os processamentos no navegador do cliente e parte do processamento no servidor da aplicação conforme a figura 6.



Figura 6. Aplicação em dispositivo móvel

Assim, o estudante poderá acessar o ambiente através do seu celular ou *tablet* e mesmo assim o sistema irá detectar e classificar suas emoções durante o processo de ensino aprendizagem.

4. Considerações Finais

[Arruda 2012] toma por base [Maturana and Garcia 1997] e considera que é a emoção que desencadeia uma ação do sujeito, ao se constituir no fundamento da razão, determinando a “estratégia em relação à construção de nossos argumentos e de nossas conversações” (ARRUDA, 2012, p. 298). Ainda, segundo Arruda (2012), o papel mediador do professor no processo de ensino-aprendizagem atribui maior ênfase ao processo dialógico (conversação), no qual são produzidos novos conhecimentos de forma conjunta entre educador e educando. Assim, cabe ao professor o papel de mediador de emoções em um processo interativo e o reconhecimento das manifestações emotivas do aprendiz pode contribuir para explorar aspectos que interferem no sucesso ou fracasso de uma aprendizagem.

4.1. Trabalhos Futuros

Como trabalho futuro será realizado um estudo e testes sobre a possibilidade da utilização de *webcam* com mais alta capacidade, como por exemplo, detecção de profundidade e o uso de aparelhos como o *Kinect* para a captura das expressões faciais com mais acurácia resultando assim em melhores classificações futuras.

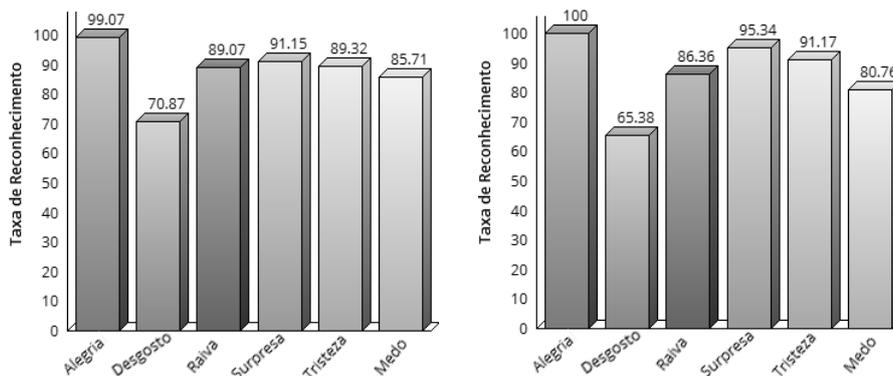
Além disso, será realizado alterações no sistema para que ele seja capaz de gravar algumas imagens das expressões faciais, teremos assim como produto final uma base de dados de emoções espontâneas de estudantes de diferentes etnias e sexo, essa base de dados pode ser utilizada para o aperfeiçoamento do algoritmo classificador.

4.2. Resultados da Validação

A Figura 7 mostra dois gráficos comparativos com resultados obtidos através dos testes efetuados com imagens e vídeos. Note que, os resultados para classificação das emoções

foram satisfatórios tanto para emoções expressas por imagens quanto por vídeos.

Uma exceção é para a emoção “desgosto” que obteve menor acurácia comparada as demais nos dois diferentes testes. Isso se deve ao fato desta emoção possuir muitas unidades de ação correlacionadas com outras emoções, gerando assim, falsos positivos durante o reconhecimento das emoções.



a) Taxa de Reconhecimento por Imagem.

b) Taxa de Reconhecimento por Vídeo.

Figura 7. Gráfico comparativo de Taxa de Reconhecimento por Imagem e Vídeo.

Os resultados parciais para os testes com imagens obtiveram, no geral, uma média de acertos de 87,53%. Para a emoção “feliz” o total de acertos foi de 99,07%, para a “tristeza” 70,87%, para a “raiva” o total de acertos foi de 89,07%, o “desgosto” obteve um total de 91,15% acertos, “surpresa” 89,32% acertos e a emoção “medo” obteve um total de acertos na margem de 85,71%. Para a emoção “desgosto” num total de 182 imagens, 4 imagens foram confundidas com a emoção “raiva”, 2 com a emoção “medo” e 47 para “tristeza”, conforme mostrado na Tabela 2, que expõe a matriz de confusão obtida para os testes com imagens, podendo-se observar os demais resultados.

Tabela 2. Matriz de confusão obtida com testes para imagens.

Total de imagem	Alegria	Tristeza	Desgosto	Surpresa	Raiva	Medo	Neutro
324	321	0	0	0	0	0	3
253	0	233	7	0	8	12	0
182	0	47	129	0	4	2	0
328	0	2	10	299	0	17	0
183	0	12	8	0	163	0	0
182	17	2	6	0	1	156	0

Os testes com vídeos obtiveram, no geral, uma média de acertos de 86,50%. Para a emoção “feliz” o total de acertos foi de 100,00%, “tristeza” 91,17% de acertos, “raiva” 86,36% de acerto, o “desgosto” obteve um total de 65,38% acertos, “surpresa” 95,34% acertos e, “medo”, o total de acertos foi de 80,76%. Para a emoção “tristeza” – exemplo, linha 2 – foram usados um total de 34 vídeos, onde foram confundidas 3 vídeos para a emoção “desgosto”, conforme mostrado na Tabela 3, que mostra a matriz de confusão obtida para os testes com vídeos, podendo observar os demais resultados.

Tabela 3. Matriz de confusão obtida com testes para vídeos.

Total de vídeos	Alegria	Tristeza	Desgosto	Surpresa	Raiva	Medo
43	43	0	0	0	0	0
34	0	31	3	0	0	0
26	0	4	17	0	4	1
43	0	0	0	41	2	0
22	0	0	0	3	19	0
26	0	0	0	2	3	21

Ao analisar os testes, é possível observar que o reconhecimento das emoções através das Unidades de Movimentos também é uma fator complexo, pois além de detectar os pontos na face é necessário tratar a presença de pontos iguais em emoções diferentes. Mas, para as outras emoções os resultados foram satisfatórios e, a partir deles é possível reconhecer com alta taxa de precisão as emoções alegria, surpresa e tristeza.

Referências

- Arruda, M. P. (2012). O paradigma emergente da educação: o professor como mediador de emoções. *ETD-Educação Temática Digital*, 14(2):290–303.
- Azcarate, Hageloh, F., S, K. V. D., and Valenti, R. (2005). Automatic facial emotion recognition.
- Barron, J. L., Fleet, D. J., and Beauchemin, S. S. (1994). Performance of optical flow techniques. *International journal of computer vision*, 12(1):43–77.
- Ekman, P. and Friesen, W. V. (1978). *Manual for the facial action coding system*. Consulting Psychologists Press.
- Kanade, T. C. and Cohn, J. (2005). Au-coded facial expression database.
- Maturana, H. R. and Garcia, F. J. V. (1997). *De máquinas e seres vivos: autopoiese; a organização do vivo*. Artes médicas.
- Moodle. Moodle. https://docs.moodle.org/28/en/About_Moodle. Acesso Jan 1, 2015.
- Russell, S. and Norvig, P. (2004). *Inteligência artificial*. CAMPUS - RJ.
- Schapire, R. E. (1990). The strength of weak learnability. *Mach. Learn.*, 5(2):197–227.
- Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154.