sensores

MDPI

Artigo

Processamento de Imagem Visual e Térmica para Facial Detecção de Marco Específico para Inferir Emoções em um Interação criança-robô

Christiane Goulart 1,*, Carlos Valadão 2,*, Denis Delisle-Rodriguez 2,3, Douglas Funayama 4, Álvaro Favarato 5, Guilherme Baldo 6, Vinícius Binotte 2, Eliete Caldeira 6 e Teodiano Bastos-Filho 1,2,6

1 Rede Nordeste de Biotecnologia (RENORBIO), Programa de Pós-Graduação em Biotecnologia, Centro de Ciências da Saúde, Universidade Federal do Espírito Santo (UFES), Av. Marechal Campos, 1468, Vitória-ES 29043-900, Brasil

2 Programa de Pós-Graduação em Engenharia Elétrica, UFES, Av. Fernando Ferrari, 514, Vitória-ES 29075-910, Brasil

3 Centro de Biofísica Médica, Universidade de Oriente, Patricio Lumumba s/n, Santiago de Cuba 90500, Cuba

4 Departamento de Engenharia de Computação, UFES, Av. Fernando Ferrari, 514, Vitória-ES 29075-910, Brasil

5 Departamento de Engenharia Mecânica, UFES, Av. Fernando Ferrari, 514, Vitória-ES 29075-910, Brasil

6 Departamento de Engenharia Elétrica, UFES, Av. Fernando Ferrari, 514, Vitória-ES 29075-910, Brasil

* Correspondência: christiane.ufes@gmail.com (CG); carlostvaladao@gmail.com (CV); Tel.: +55-27-4009-2661 (CG e CV)

Recebido: 20 de maio de 2019; Aceito: 22 de junho de 2019; Publicado: 26 de junho de 2019

Resumo: A Interação Criança-Robô (CRI) tem se tornado cada vez mais abordada em pesquisas e formulários. Este trabalho propõe um sistema para reconhecimento de emoções em crianças, registrando imagens por câmeras visuais (RGB—vermelho, verde e azul) e de imagem térmica infravermelha (IRTI). Para isso, o algoritmo Viola-Jones é usado em imagens coloridas para detectar regiões faciais de interesse (ROIs), que são transferidos para o plano da câmera termográfica multiplicando uma homografia matriz obtida através do processo de calibração do sistema de câmeras. Como novidade, propomos calcular a probabilidade de erro para cada ROI localizada em imagens térmicas, usando um quadro de referência marcado manualmente por um especialista treinado, a fim de escolher aquele ROI melhor colocado de acordo com o critérios especializados. Então, essa ROI selecionada é usada para realocar as outras ROIs, aumentando a concordância em relação às anotações do manual de referência. Depois, outros métodos para extração de características, redução de dimensionalidade através da Análise de Componentes Principais (PCA) e classificação de padrões por Análise Discriminante Linear (LDA) são aplicados para inferir emoções. Os resultados mostram que nosso abordagem para locais de ROI pode rastrear pontos de referência faciais com erros baixos significativos em relação a o algoritmo tradicional de Viola-Jones. Esses ROIs se mostraram relevantes para o reconhecimento de cinco emoções, especificamente nojo, medo, felicidade, tristeza e surpresa, com nosso sistema de reconhecimento com base em PCA e LDA alcançando precisão média (ACC) e valores Kappa de 85,75% e 81,84%, respectivamente. Como segunda etapa, o sistema de reconhecimento proposto foi treinado com um conjunto de dados de imagens, coletadas em 28 crianças com desenvolvimento típico, a fim de inferir uma das cinco emoções básicas (nojo, medo, felicidade, tristeza e surpresa) durante uma interação criança-robô. Os resultados mostram que nosso sistema pode ser integrado a um robô social para inferir as emoções da criança durante uma interação criança-robô.

Palavras-chave: Viola-Jones; reconhecimento de emoções faciais; reconhecimento de expressão facial; detecção facial; marcos faciais; imagem térmica infravermelha; matriz de homografia; robô socialmente assistido

1. Introdução

Interação Criança-Robô (CRI) é um subcampo da Interação Humano-Robô (HRI) [1], que é definida como a interação entre humanos e sistemas robóticos. Dentro das diversas possibilidades de HRI e CRI, robôs socialmente assistivos estão sendo utilizados como ferramenta de auxílio terapêutico para crianças com autismo [2,3]. Uma característica que pode melhorar essa interação é a capacidade de reconhecer emoções, que pode ser utilizada para fornecer um IRC melhor. Por exemplo, crianças com transtorno do espectro do autismo (TEA) tendem a não ter a capacidade de exibir emoções, portanto, os robôs devem confiar em medições involuntárias de sinais biológicos, como termografia da pele [4-6].

A face é uma região do corpo que apresenta alta resposta às emoções, e as alterações da impressão térmica facial podem estar ligadas à emoção da criança. Assim, esta característica pode ser um parâmetro útil a ser aplicado em um IRC, uma vez que este sinal biológico não é voluntário e não é facilmente mutável [7]. Devido a esta característica, estudos recentes estão focados na detecção facial e termografia para avaliar expressões emocionais na computação afetiva [8-10]. Além disso, é uma técnica mais confortável e discreta para avaliar emoções, pois não é necessário nenhum sensor que toque a criança, como eletrodos usados em eletroencefalografia e eletrocardiografia [8,11,12].

Um sistema convencional para reconhecimento de emoções faciais é composto por três etapas principais: detecção de face e componentes faciais, computação de várias características espaciais e temporais e classificação de emoções . Então, a primeira etapa para detecção de rosto sobre uma imagem de entrada e, consequentemente, para localizar componentes faciais (como olhos, nariz e boca) ou pontos de interesse, é uma tarefa crucial e ainda um desafio. De fato, para discriminar com precisão as emoções, é necessário aplicar métodos baseados em características geométricas ou de aparência [10,13–15], sendo este último o mais popular, devido ao seu desempenho superior [16]. Por outro lado, os Marcos Faciais (LF) devem ser usados para localizar pontos salientes de regiões faciais, como a extremidade do nariz, as extremidades das sobrancelhas e a boca [10,16].

Muitos estudos demonstraram que dividir o rosto em regiões específicas para extração de características faciais pode melhorar o desempenho durante o reconhecimento de emoções [17-29]. No entanto, essa estratégia pode ser afetada pelo alinhamento inadequado da face. Além disso, outros trabalhos baseados em aprendizado [30,31] para extração de características de regiões faciais específicas foram propostos para localizar as regiões faciais com maior contribuição para o reconhecimento de emoções. No entanto, essas abordagens são difíceis de serem estendidas como um sistema genérico, devido ao fato de que as posições e tamanhos das manchas faciais variam de acordo com os dados de treinamento.

Vale comentar que estudos com câmeras térmicas para reconhecimento de emoções têm mostrado resultados promissores, mas câmeras térmicas de baixo custo normalmente apresentam baixa resolução, dificultando a detecção precisa das regiões faciais pela aplicação de métodos convencionais como o algoritmo Viola-Jones, amplamente utilizado em imagens visuais [15,32].

Então, levantamos a hipótese de que um sistema de baixo custo para captura simultânea de câmeras visuais e térmicas pode aumentar a precisão para localizações de regiões faciais de interesse (ROIs) específicas sobre faces e, consequentemente, melhorar a extração de recursos, aumentando a discriminação de emoções.
Desta forma, consideramos uma alternativa pouco explorada, que é primeiramente aplicar o algoritmo Viola-Jones na imagem visual para localizar as ROIs desejadas, e depois transferi-la para sua imagem térmica correspondente, mas incluindo como última etapa um método para localização de ROIs correção baseada na probabilidade de erro, levando em consideração anotações manuais de um especialista treinado sobre um quadro de referência.

Assim, o objetivo deste trabalho é propor um sistema capaz de detectar ROIs faciais para cinco emoções (nojo, medo, felicidade, tristeza e surpresa) em crianças com desenvolvimento típico durante uma interação com um robô social (como estímulo afetivo). Neste estudo, nosso sistema de câmera de baixo custo permite obter pares de imagens sincronizadas para detectar ROIs na imagem visual usando o algoritmo Viola-Jones como primeiro estágio e, em seguida, transferir essas ROIs para o quadro da câmera térmica correspondente por meio de uma matriz de homografia . Como principal novidade, apresentamos aqui uma nova maneira de melhorar com precisão as localizações do ROI após a aplicação de Viola-Jones e transformação de homografia. Essa abordagem calcula a probabilidade de erro para encontrar automaticamente o ROI localizado nas imagens térmicas, que está melhor posicionada

de acordo com as anotações manuais de um especialista treinado. Este ROI de maior probabilidade (com o menor erro de localização) é usado posteriormente para realocar outros ROIs, melhorando a precisão geral. Então, melhores características de aparência podem ser extraídas, a fim de aumentar a discriminação de emoções pelo nosso sistema de reconhecimento proposto. Da mesma forma, este método pode ser estendido a outros estudos que visam localizar com precisão ROIs sobre imagens térmicas faciais, que são fisiologicamente relevantes, como descrito em [17,21], permitindo entender fenômenos ligados a comportamentos, emoções, estresse, interações humanas, entre outros. Como relevância deste trabalho, nosso sistema é capaz de detectar ROIs no rosto da criança, que tem importância neurofisiológica para o reconhecimento de emoções por meio de imagens térmicas gravadas de forma discreta. Além disso, métodos para extração de características e redução de dimensionalidade são aplicados em ROIs específicos para reconhecimento de emoções usando Análise Discriminante Linear (LDA). Como outro destaque, um conjunto de imagens visuais e térmicas é adquirido em um contexto atípico em que um robô social é usado como estímulo emocional em uma interação com crianças, a fim de testar o sistema proposto para detecção de ROIs específicas e reconhecimento de emoções. Para nosso conhecimento, esse tipo de abordagem não foi explorado em outros estudos.

Este trabalho está estruturado da seguinte forma. A seção 2 apresenta uma descrição de vários trabalhos do estado da arte. A seção 3 apresenta um sistema para aquisição de imagens, além de uma proposta baseada no algoritmo de Viola-Jones e probabilidade de erro para localização de ROI facial. Além disso, o protocolo experimental e métodos para extração de características, redução de dimensionalidade e classificação são descritos. A seção 4 apresenta os resultados experimentais sobre o método automático de posicionamento de ROI e reconhecimento de emoções das crianças durante a interação com o robô. Em seguida, a Seção 5 apresenta os achados deste trabalho e os compara com estudos anteriores, resumindo também suas principais contribuições e limitações. Por fim, a Seção 6 apresenta a Conclusão e os Trabalhos Futuros.

2. Trabalhos Relacionados

Muitas pesquisas para reconhecer a emoção facial por estratégias sem contato propuseram métodos automáticos para detecção de ROI facial e facial em imagens visuais e térmicas, pois construir uma representação facial eficaz a partir de imagens é um passo crucial para uma análise automática de ação facial bem-sucedida, a fim de reconhecer emoções faciais. Nesse campo, existe o Facial Action Coding System (FACS), que é uma taxonomia de expressões faciais humanas projetada para facilitar a anotação humana do comportamento facial [9,14,33]. Por exemplo, um total de 32 ações musculares faciais atômicas, denominadas Unidades de Ação (AUs), e 14 descritores adicionais relacionados a ações diversas são especificados, que são amplamente utilizados por métodos automáticos para localizar pontos de referência faciais e ROIs. Essas regiões são usadas por métodos baseados em características geométricas [9,10,15] e de aparência para discriminar emoções [9,14]. As representações de aparência usam informações de textura considerando o valor de intensidade dos pixels, enquanto as representações geométricas ignoram a textura e descrevem a forma explicitamente [9,14,15]. Aqui, focamos nossa revisão do estado da arte em abordagens usando apenas recursos de aparência na face alvo, que geralmente são calculados dividindo a região da face em grade regular (representação holística).
Os recursos de aparência podem ser obtidos para codificar informações de baixo ou alto nível. Por exemplo, informações de baixo nível podem ser codificadas por meio de histogramas de baixo nível computacionalmente simples e ideais para aplicações em tempo real, representações Gabor, representações orientadas a dados por aplicação de bag-of-words, entre outros. Além disso, um nível mais alto de informação pode ser codificado através da fatoração matricial não negativa (NMF) [9]. No entanto, a eficácia da extração de características para aumentar a discriminação de emoções pode ser afetada por vários fatores, como variações de postura da cabeça, variações de iluminação, registro de face, oclusões, entre outros [9].

Em [16] os autores utilizaram o classificador Haar para detecção de face, que é amplamente aplicado, devido à sua alta precisão de detecção e desempenho em tempo real [32]. Eles extraíram características de aparência da região global da face aplicando o histograma do Padrão Binário Local (LBP) que cuida de pequenas alterações da expressão facial para diferentes emoções [9,34], seguido de Análise de Componentes Principais (PCA) para redução de dimensionalidade, para melhorar a velocidade de computação em tempo real durante seis emoções (raiva, desgosto, medo, felicidade, tristeza e surpresa). Essa abordagem é personalizável de pessoa para pessoa e

alcançou uma precisão (ACC) de 97%. Vale ressaltar que, diferentemente de uma abordagem baseada em recursos globais, diferentes regiões da face têm diferentes níveis de importância para o reconhecimento de emoções [17]. Por exemplo, os olhos e a boca contêm mais informações do que a testa e a bochecha. Observe que a LBP tem sido amplamente utilizada em muitas pesquisas de reconhecimento de emoções. Consulte a Ref. [34] para um estudo abrangente sobre métodos baseados em LBP para reconhecimento de emoções.

Outro estudo [14] usou regiões específicas para extração de características de aparência dividindo toda a região da face em regiões locais específicas de domínio, usando o método de detecção de pontos de referência apresentado na Ref. [35] que usa ensemble de árvores de regressão. Esses autores usaram localizações de pontos faciais para definir um conjunto de 29 regiões da face cobrindo toda a face, que foi baseado no conhecimento especializado sobre geometria da face e contrações musculares faciais específicas da AU, como mostrado na Ref. [33]. O conjunto de árvores de regressão é usado para estimar as localizações dos pontos de referência da face diretamente de um subconjunto esparso de intensidades de pixels, alcançando desempenho em tempo super-real com previsões de alta qualidade. Da mesma forma, eles usaram o descritor LPB para extração de recursos de aparência, alcançando um ACC de 93,60% após a aplicação de Support Vector Machine (SVM) com kernel Radial Basic Function (RBF).

Na Ref. [36], um estudo comparativo de métodos para extração de características, como Kernel Discriminant Isometric Mapping (KDIsomap), PCA, Linear Discriminant Analysis (LDA), Kernel Principal Component Analysis (KPCA), Kernel Linear Discriminant Analysis (KLDA) e Kernel O Mapeamento Isométrico (KIsomap) foi realizado, obtendo o melhor desempenho (ACC de 81,59% no banco de dados JAFFE e 94,88% no banco de dados Cohn-Kanade) para o KDIsomap durante sete emoções (raiva, alegria, tristeza, neutro, surpresa, desgosto e medo), mas sem diferença significativa em comparação com outras abordagens. Aqui, os autores usaram o conhecido algoritmo Viola-Jones para detectar a face [32], que é adequado para aplicações em tempo real. Este método usa uma cascata de classificadores empregando recursos Haar-wavelet, que geralmente usam a posição do olho detectada na região da face para alinhar as outras regiões da face detectadas.

Na Ref. [37] os autores propõem o algoritmo Central Symmetric Local Gradient Coding (CS-LGC) para definir a vizinhança como uma grade 5 × 5, usando o conceito de simetria central para extrair a informação do gradiente em quatro direções (horizontal, vertical e duas direções). diagonais) para extração de recursos sobre pixels de destino mais representativos. Em seguida, eles também aplicaram o PCA para redução de dimensionalidade, seguido pelo algoritmo Extreme Learning Machine (ELM). A avaliação dessa abordagem foi realizada por meio dos bancos de dados JAFFE e Cohn-Kanade, que contêm imagens visuais em escala de cinza relacionadas às seguintes emoções: raiva, nojo, medo, felicidade, neutro, tristeza e surpresa. Precisões de 98,33% e 95,24% para Cohn-Kanade e JAFFE foram obtidas, respectivamente, sendo relativamente melhores em comparação com outros operadores para extração de características, como o LBP.

Vários estudos para reconhecimento de emoções têm sido realizados com dois tipos de câmera (uma visual e outra infravermelha), como na Ref. [20]. Esses autores propuseram um esquema de fusão aplicando PCA sobre faces térmicas e visuais para extração de recursos e k vizinhos mais próximos para reconhecer duas classes (surpresa e rindo) com ACC médio de 75%. Além disso, na Ref. [23] uma comparação para reconhecimento de emoções usando câmeras visuais e infravermelhas foi realizada e quatro métodos típicos, incluindo PCA, PCA mais LDA, Active Appearance Model (AAM) e baseado em AAM mais LDA foram usados em imagens visuais para extração de recursos, enquanto PCA e PCA mais LDA foram aplicados em imagens térmicas infravermelhas usando quatro ROIs (testa, nariz, boca e bochechas). Esses autores usaram k-vizinhos mais próximos para reconhecer seis emoções (tristeza, raiva, surpresa, medo, felicidade e nojo).
Vale ressaltar que as localizações dos olhos sobre a térmica foram realizadas manualmente por esses autores, que posteriormente foram utilizadas para localizar as quatro ROIs mencionadas. Na Ref. [22] foi abordada uma abordagem interessante usando os dois tipos de câmera, incluindo o uso de óculos, que são opacos para a câmera térmica, mas visíveis para a câmera visual.

Outro trabalho interessante mostra que câmeras infravermelhas e visuais podem ser combinadas em um sistema de sensores multimodal para reconhecer o medo [24], por meio de sinais de eletroencefalograma (EEG), taxa de piscar de olhos e temperatura facial enquanto o usuário assiste a um filme de terror. Um algoritmo Adaptive Boosting (AdaBoost) foi usado para detectar a região da face, e uma transformação geométrica para fazer a

foram utilizadas as coordenadas das duas imagens (luz visível e térmica) coincidentes. Da mesma forma, outro estudo foi realizado em pacientes com transtorno de estresse pós-traumático (TEPT) para inferir o medo por meio de imagens visuais e térmicas [38]. Na Ref. [39], foi proposto um algoritmo para determinação automática do centro da cabeça em termogramas, que demonstrou ser sensível à rotação ou posição da cabeça.
Na Ref. [40], os autores propuseram uma extração de recursos locais e globais não supervisionados para reconhecimento de emoções faciais por meio de imagens térmicas. Para tanto, utilizaram um limiar bimodal para localizar a face para extração de características por PCA, após aplicar um método baseado em agrupamento para detectar pontos de interesse; para a classificação da expressão facial, foi utilizado um Comitê de Máquinas de Vetores de Suporte.
Na Ref. [17], a face foi extraída em imagens térmicas após a aplicação de filtros medianos e gaussianos com posterior binarização para converter a imagem em escala de cinza em preto e branco puro e remover pequenos conjuntos de pixels não conectados para melhorar a qualidade da imagem. Em seguida, as características de aparência foram extraídas em ROIs definidas sobre as imagens térmicas, seguidas de Análise de Componentes de Vizinhança Rápida (FNCA) e LDA para seleção de características e reconhecimento de cinco emoções, respectivamente.

Mais detalhes sobre diferentes métodos para extração de características, redução de dimensionalidade, seleção de características e classificação podem ser revistos em alguns estudos [37] e também em revisões extensas, como nas Refs. [9,10].

A próxima seção apresenta nosso sistema proposto para o reconhecimento de cinco emoções, que permite localizar com precisão ROIs faciais sobre imagens térmicas, melhorando a extração de recursos de aparência.

3. Materiais e Métodos

3.1. Procedimento experimental

Participaram deste estudo 17 crianças com desenvolvimento típico, 9 meninos e 8 meninas (com idades entre 8 e 12 anos), recrutadas em escolas de ensino fundamental de Vitória-Brasil. Todos tiveram a autorização dos pais, por meio da assinatura do Termo de Consentimento Livre e Esclarecido. Além disso, as crianças assinaram um Termo de Assentimento, informando o desejo de participar. Este estudo foi aprovado pelo Comitê de Ética da Universidade Federal do Espírito Santo (UFES)/Brasil, sob o número 1.121.638.
Os experimentos foram conduzidos em uma sala dentro do ambiente escolar das crianças, onde a temperatura ambiente foi mantida entre 20 ÿC e 24 ÿC, utilizando uma intensidade luminosa constante, como feito pela Ref. [39].

Um robô social móvel (ver Figura 1b), denominado N-MARIA (New-Mobile Autonomous Robot for Interaction with Autistics), construído na UFES/Brasil para auxiliar crianças durante a reabilitação do relacionamento social, foi utilizado em nossa pesquisa. Este robô anexou um sistema de câmera para registrar imagens faciais durante a interação com as crianças. Mais detalhes sobre N-MARIA são fornecidos na Seção 3.2.1.

O experimento foi conduzido em três fases, conforme segue. Primeiramente, o N-MARIA foi inicialmente coberto na sala com um lençol preto, exceto seu sistema de câmera acoplada, que foi acionado para registrar imagens visuais e térmicas da vista frontal com taxa de amostragem de 2 fps, para posterior processamento. Em seguida, a criança foi convidada a entrar na sala e sentar-se confortavelmente para explicações sobre as atividades gerais relacionadas ao experimento, sendo condicionada a um estado relaxado por um período mínimo de 10 min, a fim de adaptar seu corpo às a temperatura da sala, permitindo que a temperatura da sua pele se estabilize para registros de linha de base, de acordo com estudos semelhantes realizados nas Refs. [21,41]. Terminado o período de relaxamento, a criança foi colocada em frente ao robô coberto a cerca de 70 cm de distância do mesmo, permanecendo em pé. Imediatamente, foram realizadas gravações do rosto da criança pelo sistema de câmeras por um período de um minuto com o robô coberto, um minuto com o robô descoberto e três minutos de interação com o robô. Após, a criança passou dois minutos respondendo a um questionário sobre o experimento.

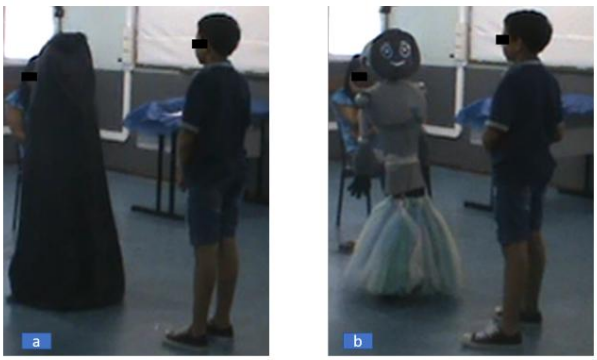

Figura 1. Configuração experimental mostrando a interação criança-robô. **(a)** Antes de mostrar o robô; **(b)** Depois de apresentá-lo.

A primeira parte da gravação (robô coberto) corresponde à etapa experimental, chamada de Linha de Base, enquanto a próxima etapa, apresentando o robô descoberto, é chamada de Teste. Antes que o robô fosse descoberto, a criança foi solicitada a olhar permanentemente para frente sem movimentos faciais bruscos ou tocar o rosto, evitando qualquer obstrução facial durante as gravações de vídeo.

Após a retirada do lençol preto que cobria o robô, iniciou-se o primeiro diálogo (auto-apresentação) do robô. Além da autoapresentação à criança, diálogos pontuais durante o experimento foram relacionados a perguntas, reforço positivo e convites. Na interação com a criança, que durou dois minutos, a criança foi estimulada a fazer comunicação e interação tátil com o robô. Ao final do experimento, a criança foi novamente convidada a sentar e responder uma entrevista estruturada sobre seus sentimentos antes e depois de ver o robô, e também sobre a estrutura do robô (se a criança gostou, o que ela gostou mais). , e o que a criança mudaria nela).

3.2. Reconhecimento de emoções sem contato

A Figura 2 mostra o sistema de reconhecimento de emoções sem contato proposto, composto das quatro etapas a seguir: (a) calibração da câmera; (b) aquisição de imagens e detecção automática de ROI; (c) substituição do ROI; (d) extração de características seguida da redução da dimensionalidade e classificação das emoções.

A Figura 2a mostra uma primeira etapa para calibrar o sistema de câmeras obtendo uma matriz de homografia para mapear os pixels da imagem da câmera visual na imagem da câmera térmica, considerando a posição fixa relativa entre as duas câmeras. Além disso, outro processo é realizado para obter um quadro que contém ruído intrínseco do sensor infravermelho, que é posteriormente utilizado em uma segunda etapa (Figura 2b) para remover o ruído do sensor (inerente à câmera) sobre a imagem térmica atual capturada. Nesta segunda etapa, o processo de aquisição de imagem é realizado a partir de imagens síncronas de câmeras visuais e infravermelhas, que são pré-processadas para aprimorar a detecção automática de ROIs faciais aplicando o algoritmo Viola-Jones na imagem visual. Em seguida, as ROIs colocadas na imagem visual são projetadas na imagem térmica usando a matriz de homografia. Como terceiro estágio, anotações manuais por um especialista treinado sobre um quadro de referência são usadas para realocar com precisão os ROIs aplicando nossa abordagem com base em erros de probabilidade, como mostrado na Figura 2c. Em seguida, vetores de características relacionados a variações térmicas são computados nas ROIs detectadas, e depois reduzidos pela aplicação de PCA para redução de dimensionalidade para reconhecimento de cinco emoções em um último estágio por LDA. Mais detalhes sobre o sistema de reconhecimento proposto são fornecidos nas próximas subseções.

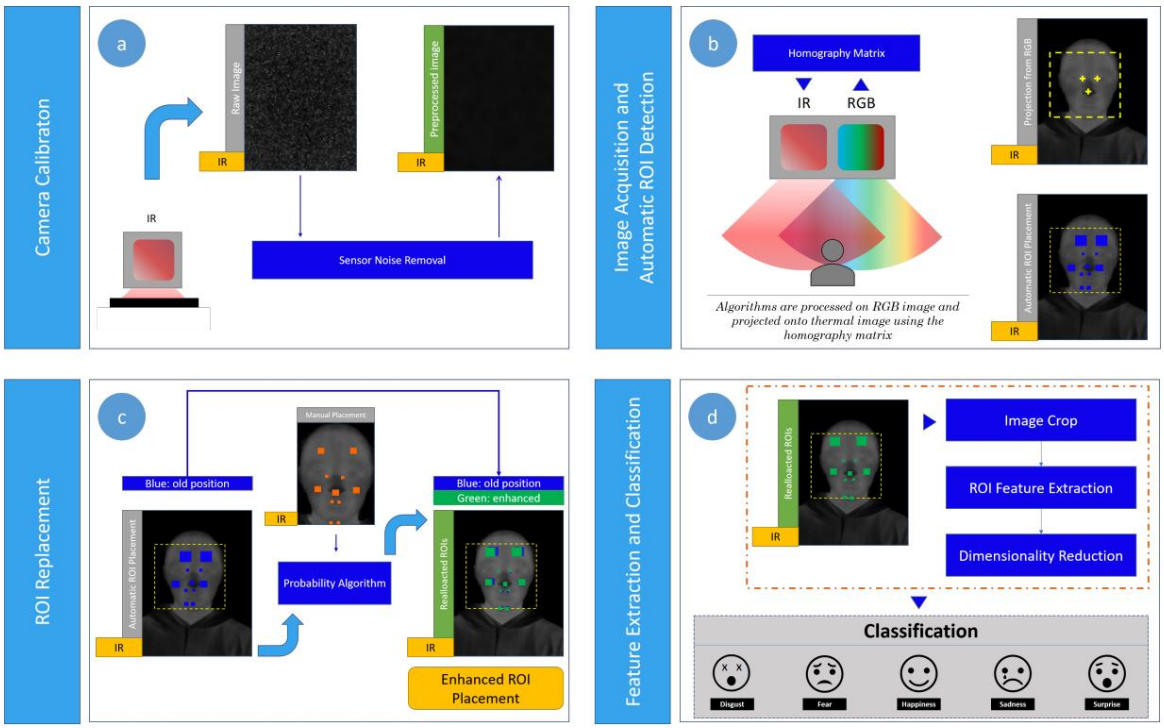

Figura 2. Visão geral do sistema proposto para reconhecimento de emoções durante uma interação criança-robô. **(a)** Calibração da câmera; **(b)** aquisição de imagem e detecção automática da região de interesse (ROI); **(c)** substituição do ROI ; **(d)** extração de características seguida da redução da dimensionalidade e classificação das emoções.

3.2.1. Sistema de câmeras e N-MARIA

O sistema de câmeras composto por câmeras visuais e térmicas foi acoplado à cabeça do robô social, conforme mostrado na Figura 3. Essas duas câmeras foram fixadas de modo que ambas tivessem aproximadamente o mesmo campo visual. Para capturar as variações térmicas, uma câmera de baixo custo (Therm-App) que possui, resolução espacial de 384 × 288 ppi, taxa de quadros de 8,7 Hz e sensibilidade à temperatura <0,07 ÿC. foi usado.
A normalização das imagens térmicas adquiridas em escala de cinza consistiu em uma taxa de brilho variando de 0 a 255, onde pixels mais escuros correspondem a temperaturas mais baixas e pixels mais claros correspondem a temperaturas mais altas. Além disso, foi utilizada uma Webcam HD C270 (Logitech) para obtenção de imagens visuais no formato RGB, com resolução de 1,2 MP.

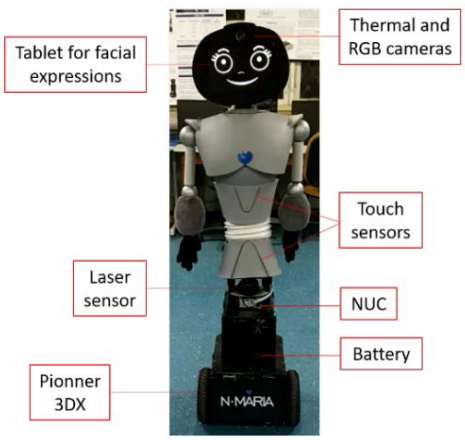

Figura 3. N-MARIA (Novo Robô Móvel para Interação com Autistas) desenvolvido na Universidade Federal do Espírito Santo (UFES)/Brasil.

O robô foi construído com 1,41 m de altura, considerando a altura padrão de crianças de 9 a 10 anos. Além disso, materiais macios e maleáveis foram usados na estrutura do robô para proteção de ambas as crianças.

e dispositivos robóticos internos. A plataforma móvel Pioneer 3-DX foi responsável pela locomoção, um sensor a laser 360ÿ foi utilizado para localizar a criança no ambiente e um tablet foi utilizado como face do robô para exibir sete expressões faciais dinâmicas durante a interação robô-criança. Essas expressões também poderiam ser controladas remotamente através de outro tablet pelo terapeuta, que também poderia controlar o comportamento do robô, as expressões e os diálogos emitidos pelos falantes.

3.2.2. Calibração da câmera

A calibração da câmera é feita através de uma aquisição síncrona entre imagens visuais e térmicas , e utilizando um tabuleiro de xadrez construído com alumínio e fita isolante posicionado em diversos ângulos possíveis. Em seguida, as imagens obtidas são processadas com o software de calibração OpenCV [42], que utiliza a Transformada Linear Direta (DLT) para retornar uma matriz de homografia [43], permitindo a transformação de pontos da imagem visual para a imagem térmica de forma robusta [32].]. Vale ressaltar que não existe uma matriz de homografia que corresponda exatamente aos pontos em todas as regiões da face (pois não estão no mesmo plano), mas a matriz obtida por DLT é utilizada como uma aproximação eficiente.

Além disso, outro procedimento para remover o ruído térmico intrínseco dos sensores infravermelhos é realizado [44], o que aumenta a qualidade das imagens térmicas corrigindo deslocamentos indesejáveis. Para isso, é registrada uma referência de um objeto com temperatura corporal uniforme cobrindo o campo visual da câmera térmica, que contém o ruído intrínseco do sensor infravermelho. Assim, esperava-se ter um quadro com o mesmo brilho para todos os pixels, porém, isso não ocorreu. Assim, o quadro com o ruído intrínseco do sensor foi utilizado na etapa de pré-processamento para eliminar o ruído térmico, conforme descrito na próxima seção.

3.2.3. Aquisição e pré-processamento de imagens

A câmera térmica tem capacidade máxima de aquisição de 8,7 fps enquanto a câmera visual tem capacidade máxima de 30 fps. Assim, para obter consistência temporal, tanto imagens visuais quanto térmicas foram gravadas simultaneamente com uma taxa de amostragem de 2 fps, o que foi adequado para nossos propósitos.

Durante a aquisição, o quadro com o ruído intrínseco do sensor obtido na etapa de Calibração foi utilizado para remover o ruído térmico intrínseco da imagem atual adquirida subtraindo pixel a pixel. Finalmente, um filtro mediano foi usado para reduzir o ruído de sal e pimenta da imagem térmica.

3.2.4. Detecção de ponto de referência de rosto

Um método automático foi proposto aqui para detecção de marcos de face em um determinado conjunto de quadros **(I** = {i1,i2, . . . ,ib , . . . ,iB}), tomando como referência ROIs anotadas por um especialista treinado no quadro iA(b = A), conforme mostrado na Figura 4. É possível observar que essas anotações manuais foram localizadas em onze ROIs **(RA** = **{R }}** de imagens térmicas, levando em consideração $a^{1}_{A}, ^{2}_{A}, \ldots, ^{k}_{A}, \ldots,$ **R** $^{11}_{UMA}$ relevância desses ROIs em outros estudos para reconhecimento de emoções faciais [17,21]. Aqui, os tamanhos de ROI facial foram calculados da mesma forma que nas Refs. [17,18,21], utilizando a largura da ROI da cabeça e as seguintes proporções definidas [18]: 6,49% para nariz, 14,28% para testa, 3,24% para região periorbitária, 9,74% para bochecha, 3,24% para região perinasal , e 5,19% para o queixo [17].

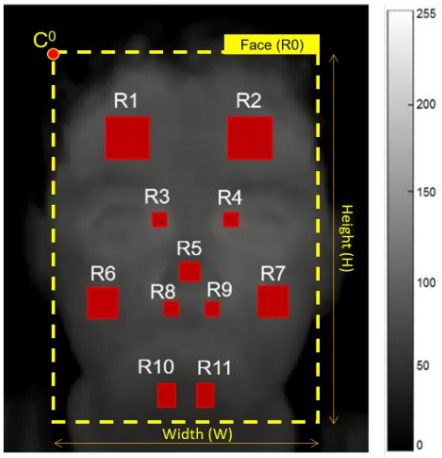

Figura 4. ROIs faciais. **R** 1, lado direito da testa; **R** 2, lado esquerdo da testa; **R**3, lado periorbitário direito; **R**4, deixei lado periorbital; lado **R** ;5, ponta do nariz; **R**6, bochecha direita; **R**7, bochecha esquerda; **R**8, lado perinasal direito; **R**9, perinasal esquerdo **R** 10, lado direito do queixo; **R** 11, lado esquerdo do queixo.

3.2.5. Detecção automática de ROI

Imagens infravermelhas são mais borradas do que imagens coloridas [23], portanto a detecção de ROI sobre imagens térmicas de câmeras de baixo custo é um desafio. Por esta razão, o conhecido algoritmo Viola-Jones [32] foi usado em imagens coloridas para detecção de cabeça e outras regiões faciais, como nariz e olhos [17,21,32].
Em seguida, essas regiões detectadas iniciais foram usadas como referências para localizar automaticamente onze ROIs na face (consulte a Tabela 1), ou seja, o nariz, ambos os lados da testa, bochechas, queixo, área periorbital (próximo aos olhos) e área perinasal (parte inferior do nariz).

Tabela 1. ROIs de referência usadas para localizar pontos de referência de face em um quadro ib .

ROIs de referência	ROIs localizados
Cabeça	**R** 1 2 **R** 11 **R** b , b , b , b 1 2 6 7 8 9 **R**
Olhos	**R R R R R** b , b , b , b , b 3 4 5 10 11 **R**
Nariz	**R R R R** b , b , b , b **R**

R 1 b , lado direito da testa; **R** 2 b , lado esquerdo da testa; **R** 3 b , lado periorbitário direito; **R** 4 b , lado periorbitário esquerdo; **R** 5 b , ponta de 6 nariz; **R** b , bochecha direita; **R**7 b , bochecha esquerda; **R** 8 b , lado perinasal direito; **R** 9 b , lado perinasal esquerdo; **R** 10 b , lado direito do queixo; **R** 11 b , lado esquerdo do queixo.

Em nosso estudo, os tamanhos de ROI facial também foram calculados usando a largura da cabeça e as proporções acima mencionadas [17,18,21]. Além disso, as ROIs faciais foram posicionadas espacialmente, tomando como referência a anotação do especialista. Em seguida, as ROIs faciais correspondentes foram projetadas na imagem térmica através da transformação mencionada usando uma matriz de homografia (consulte a Seção 3.2.2), conforme mostrado na Figura 4. Aqui, o conjunto ROI de uma moldura térmica b é definido por $R_b = \{\mathbf{R}^0_b, \mathbf{R}^1_b, \mathbf{R}^2_b, \ldots, \mathbf{R}^{11}_b\}$, sendo $\mathbf{R}^0_b$ a ROI principal e $\mathbf{R}^k_b$ k para k = 1 a 11 as ROIs faciais. Perceber k que $\mathbf{R}_b$ é descrito por vários pixels Rij para um intervalo de 0 a 255 (escala de cinza de 8 bits).

A Figura 2b,c mostra nossa proposta para localizar com precisão ROIs faciais, formadas pelas seguintes duas etapas: (1) detecção automática de ROI e (2) correção de posicionamento de ROI.

3.2.6. Correção de localização do ROI

Um novo método é proposto aqui para corrigir com precisão as ROIs detectadas pelo algoritmo Viola-Jones, levando em consideração todas as posições de ROIs pré-definidas, que foram anotadas manualmente em um primeiro quadro por um especialista treinado.

Deixe $\mathbf{R}_{bk}^{k}$ ser um ROI detectado automaticamente sobre o quadro térmico ib que apresenta uma coordenada $\mathbf{C}_{b}^{k} = (C_{Cbx}^{k}, {}_{k}by)$ no canto superior esquerdo, que corresponde a $\mathbf{R}_{UMA}^{k}$ (ROI anotado sobre iA) com coordenada $\mathbf{C}_{A} = (C_{Machado,}^{kk} C_{Ay})$ no canto superior esquerdo também. Então, duas probabilidades p $_{bx}^{k}$ e p $_{p}^{k}$ de p são calculados em relação a **xey ,** respectivamente. Eles assumem valores mais próximos de 1 se a localização $\mathbf{R}_{b}^{k}$ concorda muito com a anotação manual do especialista treinado, que são mostradas nas Equações (1) e (2). Observe que $\mathbf{R}_{b}^{k}$ é obtido automaticamente pela aplicação do algoritmo Viola-Jones, fixando proporções (ver Seção 3.2.4).

$$kpbx\_ = \frac{\exp\left(ÿ\left|\frac{C_{Cbx\_}^{k}}{Wb} \frac{C_{Machado}^{k}}{WA}\right|\right)}{\ddot{y}_{i=1}^{11} \exp\left(ÿ\left|\frac{C_{bx}^{i}}{Wb} \frac{C_{Machado}^{eu}}{WA}\right|\right)}, \quad (1)$$

$$p_{por}^{k} = \frac{\exp\left(ÿ\left|\frac{c_{por}^{k}}{Hb} \frac{C_{Sim}^{k}}{HA}\right|\right)}{por\ ÿ_{i=1}^{11} \exp\left(ÿ\left|\frac{C}{Hb} \frac{C^{eu}}{AH}\right|\right)}, \quad (2)$$

$$p_{b}^{k} = \min np_{bx,\ p}^{k}{}_{tchau}, \quad (3)$$

onde k refere-se ao ROI facial atual para análise, tomando valores de 1 a 11; Wb e Hb são a largura e a altura de **R** (head ROI para iA); p_{ib}^{0} (ROI principal para ib), respectivamente; WA e HA são a largura e a altura de $\mathbf{R}_{UMA}^{0}$ anotação do especialista treinado, em relação aos eixos xey, respectivamente. re fk_{bx}^{k} é denotado como **C) de** e $\mathbf{R}_{b}^{k}$ são as probabilidades de que **R** por foram localizados corretamente no ib em relação ao

Finalmente, $\mathbf{R}_{b}^{k}$ (para o qual $\mathbf{C}_{b}$ b $_{b}^{k}$ é selecionado como referência para corrigir a localização das outras ROIs, usando as Equações (4) e (5). Vale ressaltar que a notação **C** é usada para $_{UMA}$ o quadro anotado iA.

$$C_{Cbx\_}^{k^{0}} = \frac{C_{bx}^{ref} + (C_{Machado}^{k} - C_{Machado}^{ref})}{C}, \quad (4)$$

$$C_{por}^{k^{0}} = \frac{C^{ref\ ref} + (C_{Sim}^{k} ÿ\ C_{Sim}^{por})}{H}, \quad (5)$$

onde **C** $b^{k^{0}} = (C_{Cbx,}^{k^{0}} {}^{k^{0}}by)$ é a coordenada do canto superior esquerdo para $\mathbf{R}_{b}^{k}$ mudou-se.

3.2.7. Extração de recursos

Dado um quadro térmico ib formado por um conjunto de ROIs, Rb = **{R** $_{b}^{1}, _{b}^{2} \ldots, _{b}^{k} \ldots, _{b}^{K}$**}**, sendo K = 11 o número total de ROIs, é possível extrair de Rb um vetor de características Fb = **{f** $_{b}^{1}, f_{b}^{k} \ldots, f_{b}^{K}$**}** de **R** que descrevem um padrão relacionado a uma emoção, sendo $f_{b}^{k} = [f_{b1}^{k}, f_{b2}^{k}, \ldots, f_{b14}^{k}]$ os descritores de características computados com [17]. Tabela 2 é o valor médio de **R** $_{b}^{k}$ é o ROI atual para extração de recursos, **R** $_{b}^{k}$ e $fb(c_{+7)}^{k}$ para c iguais de 1 a 7 são outras sete características que correspondem à diferença de características computadas ao longo de quadros consecutivos. Portanto, temos onze ROIs (veja a Figura 4) e 14 recursos para cada um deles. Isso fornece um conjunto de 154 recursos por quadro.

Tabela 2. Recursos computados em cada ROI.

Características	Equações
1. Valor médio de todo o ROI	$f_{bc}^{k} = \overline{R^{k}} = \frac{1}{m \cdot n} \sum_{i=1}^{m} \sum_{j=1}^{n} Rij$, c = 1
k 2. Variação de todo o ROI, organizado em um vetor f	$= \sigma_{bc}^{2k} = \frac{1}{(m \cdot n) - 1} \sum_{i=1}^{m} \sum_{j=1}^{n} Rij - \overline{R_{b}^{k}}^{2}$, c = 2
3. Mediana de todo o ROI, organizado como um vetor	$f_{bc}^{k} = \text{mediana}(R_{bj}^{k}{}_{j=1})$, c = 3
4. Média dos valores de variação nas linhas	$k f_{bc}^{k} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{n-1} \sum_{j=1}^{n} Rij - \overline{R_{bi}^{k}}^{2}$, c = 4
5. Média dos valores medianos nas linhas	$f_{bc}^{k} = \frac{1}{m=1} \sum \text{mediana}(R_{bi}^{k})$, c = 5
6. Média dos valores de variação nas colunas	$f_{bc}^{k} = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{m-1} \sum_{i=1}^{m} Rij - \overline{R_{bj}^{k}}^{2}$, c = 6
7. Média dos valores medianos nas colunas	$f_{bc}^{k} = \frac{1}{n} \sum_{j=1}^{n} \text{mediana}(R_{bj}^{k})$, c = 7
8. Diferença de cada item em quadros consecutivos	$f_{bc}^{k} = f_{b(c+7)}^{k} - f_{bc}^{k}$ = 1, 2, . . . , 7 (b−1)c ,

3.2.8. Redução de Dimensionalidade e Classificação de Emoções

Seja **T** = (F1, y1),(F2, y2), . . . ,**(Fb ,** yb), . . . ,**(Fn,** yn) seja o conjunto de treinamento, onde n é o número de amostras e Fi é um vetor de características d-dimensional com rótulo de classe yb − 1, 2, . . . , c. Considere SVD (Singular Value Decomposition [9,16,20,23,36] é aplicado em Fi para obter os coeficientes de componentes principais, que são usados em conjuntos de treinamento e validação para reduzir o conjunto de 154 características, em para permitir um reconhecimento de emoções robusto e rápido. Como vantagens, o PCA é pouco sensível a diferentes conjuntos de treinamento e pode superar outros métodos como LDA quando o conjunto de treinamento é pequeno [45]. Este método tem sido utilizado com sucesso em muitos estudos para representar, em subespaços de menor dimensão, os vetores de características de alta dimensão, que são obtidos através da aplicação de métodos baseados em aparência [16,23]. Vale ressaltar que antes de aplicar o PCA em nosso estudo, os vetores de características do conjunto de treinamento foram normalizados usando valores de média e desvio padrão como referência. Em seguida, a validação foi normalizada utilizando os mesmos valores de referência (média e desvio padrão) obtidos do conjunto de treinamento.

Alguns classificadores, como LDA [17,23,46] e Análise Discriminante Quadrática (QDA) [12,47] , aplicando matrizes de covariância completa e diagonal, além de outros três classificadores, como discriminação de Mahalanobis [12,48], Naives Bayes [49], e Linear Support Vector Machine (LSVM) [12,14,50] são usados em nosso estudo para atribuir objetos a uma das várias classes de emoção com base em um conjunto de recursos.

3.3. Avaliação estatística

A partir das imagens registradas nos dois momentos do experimento (Base e Teste), um conjunto de 220 quadros de termografia selecionados aleatoriamente de 11 crianças foi anotado por um especialista treinado, selecionando as ROIs definidas na Figura 4. Essas imagens anotadas foram usadas como referência para avaliar , através de distâncias euclidianas (veja a Equação (6)), a exatidão e precisão de ambos Viola-Jones sem realocação ROI e Viola-Jones aplicando nosso algoritmo de realocação ROI.

$$D = \sqrt{(Ax - Mx)^2 + (Ay - Meu)^2}, \quad (6)$$

onde (Ax, Ay) é a coordenada obtida pelo método automático e (Mx, My) é a coordenada obtida pelo método manual. Quando D está próximo de zero, isso significa alta precisão.

A análise estatística utilizada para comparação entre as duas abordagens para cada ROI foi a Teste de classificação assinada de Wilcoxon para mediana zero.

Para avaliar nosso sistema proposto para reconhecimento de emoções, um banco de dados publicado (disponível nas informações de apoio de [17] no site —Disponível em https://journals.plos.org/plosone/ artigo?id=10.1371/journal.pone.0212928) foi utilizado, que é formado por vetores de características rotulados como as cinco emoções a seguir: nojo, medo, alegria, tristeza e surpresa. Este banco de dados foi coletado em 28 crianças com desenvolvimento típico (idade: 7-11 anos) por uma câmera térmica infravermelha [17]. Vale ressaltar que esse banco de dados também foi criado com crianças de faixa etária semelhante, usando a mesma câmera térmica e conjunto de recursos (um total de 154 recursos) descritos em nosso estudo (consulte as Seções 3.1, 3.2.2 e 3.2.7), e calcular esse conjunto de recursos sobre as ROIs definidas na Figura 4. Observe que os locais corretos dessas ROIs foram inspecionados visualmente por um especialista treinado. Por esta razão, foi possível comparar o sistema de reconhecimento usando um dos seguintes métodos: PCA para redução de dimensionalidade e Fast Neighbor Component Analysis (FNCA) [17] para seleção de características. Aqui, os conjuntos de treinamento e validação foram escolhidos para várias execuções de validação cruzada (kf old = 3), e métricas como precisão (ACC), Kappa, taxa de verdadeiros positivos (TPR) e taxa de falsos positivos (FPR) foram usadas [51].

Por outro lado, este banco de dados publicado foi usado para treinar nosso sistema proposto baseado em PCA, mas apenas usando dados coletados de crianças que apresentaram ACC superior a 85% durante o reconhecimento de emoção [17]. Então, nosso sistema treinado foi usado para inferir a emoção das crianças durante nosso protocolo experimental descrito na Seção 3.1.

4. Resultados

4.1. Localização automática do ROI

O desempenho do método proposto usando critérios de especialistas treinados e probabilidade de erro para detectar pontos de referência faciais com precisão foi validado em 11 crianças por meio de dois conjuntos de 220 quadros térmicos anotados pelo especialista treinado, cada um obtido para as duas condições a seguir: (1) Linha de base, (2) Teste (consulte a Seção 3.1). A Figura 5a,b mostra que nossa proposta melhorou significativamente os posicionamentos de ROI em relação aos critérios de especialistas treinados. Para ambas as condições, o método proposto concordou (erros menores que 10 pixels) com o especialista treinado, embora as crianças tendam a fazer movimentos faciais abruptos para a segunda condição, pois podem ter sido surpreendidas pelo robô descoberto. No entanto, o algoritmo de Viola-Jones, não auxiliado pela probabilidade de erro, discordou significativamente do especialista treinado, conforme mostrado na Figura 5a,b. Observe que Viola-Jones apresentou o maior erro localizando 10 e **R** 11, que corresponde aos lados direito e esquerdo do ambos **R** queixo, respectivamente. Esses erros indesejáveis podem ter sido causados por movimentos da boca durante a fala ou por expressões faciais, como felicidade e surpresa. Como destaque, nossa proposta leva em consideração o ROI melhor localizado, **R** 5 para Linha de Base e **R** 3 para Teste, o que reduziu ao máximo o erro de localização, como mostrado 1 ou **R** na Figura 5a,b, respectivamente. Vale ressaltar que nossa abordagem para substituições de ROI usa o erro de especialista treinado) qual ROI melhor localizada [illegible] aplicando Viola-Jones. Então, esta ROI selecionada é usada para realocar as outras ROIs vizinhas. Portanto, o primeiro estágio usando Viola-Jones para detectar os três ROIs iniciais é bastante decisivo.

A Figura 6 mostra, para uma criança, o posicionamento do ROI usando o algoritmo automatizado Viola-Jones (verde), o algoritmo recalculado (azul) e o posicionamento manual (amarelo). Na Ref. [17,21], os autores demonstraram que as regiões da face (ver Figura 4) ligadas ao conjunto de ramos e sub-ramos de vasos que inervam a face são determinantes para o estudo das emoções, pois as variações de temperatura da pele nessas regiões podem ser mensuradas pelo IRTI. A Figura 6 mostra que nossa proposta baseada em erro de probabilidade pode ser usada para supervisionar métodos automáticos para localização de ROIs, a fim de melhorar a extração de características de aparência sobre ROIs detectadas e, consequentemente, aumentar o desempenho durante o reconhecimento de emoções.

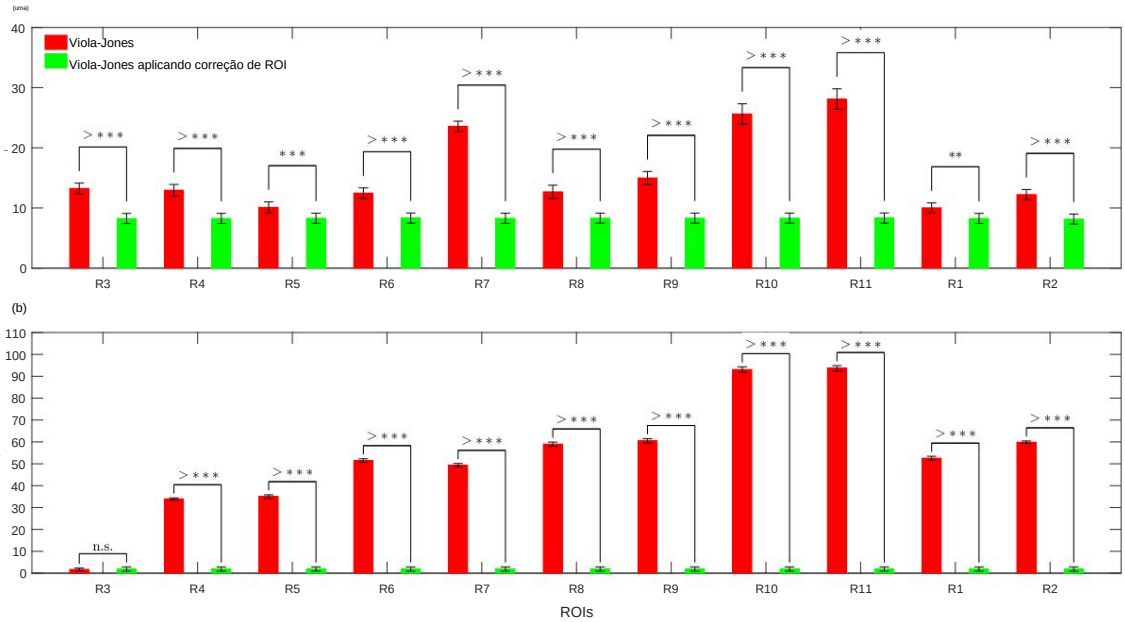

Figura 5. Comparação entre Viola-Jones (sem aplicar realocações de ROI) e Viola-Jones aplicando realocações de ROI, calculando a média e o erro padrão por ROIs: **(a)** análise da Linha de Base; **(b)** análise do Teste. ns significa nenhuma diferença significativa ($p > 0,05$), enquanto * ($p < 0,05$), ** ($p < 0,01$), *** ($p < 0,001$) e >*** ($p < 0,0001$) indicam diferença significativa. **R** ^{1}, lado direito da testa; **R** ^{2}, deixei lado da testa; **R** ^{3}, lado periorbitário direito; **R** ^{4}, lado periorbitário esquerdo; **R**^{5}, ponta do nariz; **R**^{6}, bochecha direita; **R**^{7}, deixei bochecha; **R**^{8}, lado perinasal direito; **R** ^{9}, lado perinasal esquerdo; **R** 10, lado direito do queixo; **R** 11, lado esquerdo do queixo.

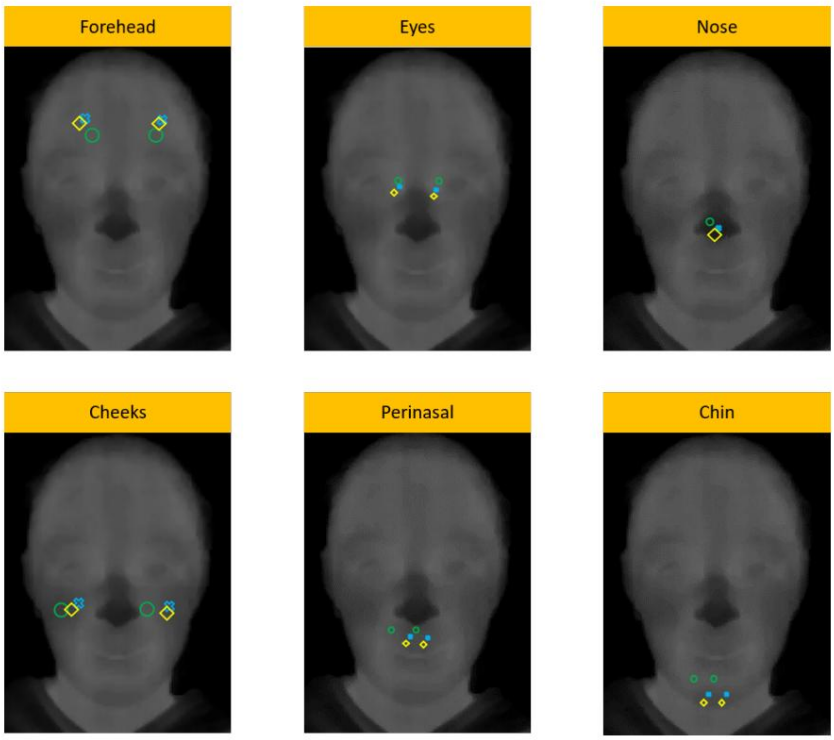

Figura 6. Comparação entre posicionamento manual (em amarelo, usado como referência), algoritmo Viola-Jones (verde) e Viola-Jones com nosso algoritmo de substituição (azul). Tal como mostrado, nossa substituição algoritmo obteve melhores resultados do que usando apenas o algoritmo Viola-Jones.

4.2. Reconhecimento de Emoções

Um banco de dados com 28 crianças com desenvolvimento típico foi utilizado para analisar o desempenho do sistema proposto para o reconhecimento de cinco emoções [17]. Esse banco de dados contém vetores de recursos de aparência que foram calculados em onze ROIs de face (consulte a Figura 4), cujos posicionamentos corretos foram cuidadosamente verificados por um especialista treinado. Então, a eficácia do PCA mais outros classificadores para a classificação de cinco emoções foi avaliada aqui, sendo o PCA uma transformação adaptativa popular que sob condições controladas de pose e imagem pode ser útil para capturar características de expressões de forma eficiente [9]. A Figura 7a mostra o desempenho médio aplicando PCA para diferentes componentes principais mais classificador LDA baseado na matriz de covariância completa, sendo 60 componentes suficientes para o reconhecimento de cinco emoções. Observe que foi alcançada diferença não significativa ($p < 0,05$) utilizando mais de 60 componentes. Da mesma forma, diferença não significativa foi obtida usando PCA com 60 componentes mais outros classificadores, como LDA usando matrizes de covariância completa (Linear) ou diagonal (Diag Linear) e SVM Linear, conforme mostrado na Figura 7b. Por exemplo, LDA usando matrizes de covariância total e diagonal alcançou valores Kappa de 81,84 ± 1,72% e 77,99 ± 1,74%, respectivamente, e 78,58 ± 1,83% para SVM Linear. O LDA baseado em matriz de covariância completa tem baixo custo computacional e pode ser facilmente incorporado ao hardware N-MARIA para reconhecimento de emoções on-line durante a interação criança-robô. Além disso, PCA e LDA têm sido utilizados com sucesso em outros estudos semelhantes [16,17,20,23,36], alcançando resultados promissores. Em seguida, selecionamos PCA com 60 componentes principais mais LDA com base na matriz de covariância completa como a melhor configuração para o reco

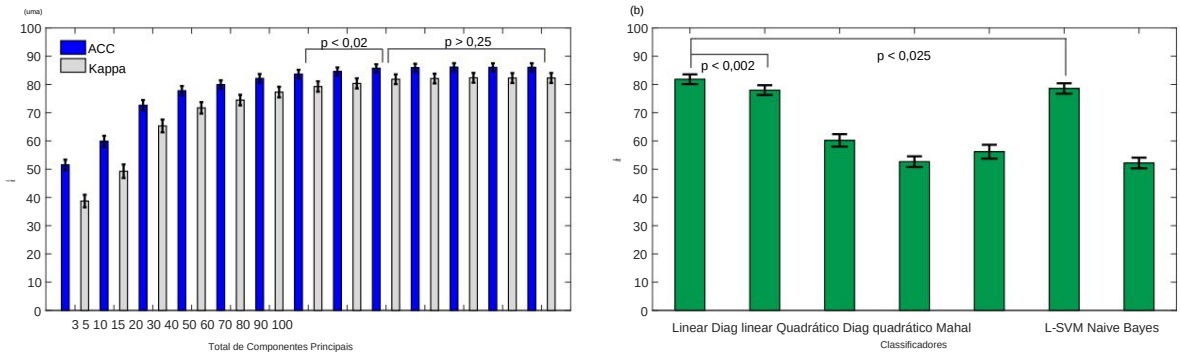

Figura 7. Desempenho da classificação de cinco emoções aplicando análise de componentes principais (PCA) para redução de dimensionalidade. **(a)** Precisão média e Kappa obtidos pela aplicação de PCA para diferentes componentes principais mais Análise Discriminante Linear para classificação; **(b)** Kappa médio alcançado para 60 componentes principais mais outros classificadores.

Consequentemente, uma comparação entre PCA usando 60 componentes e FNCA foi realizada para reconhecer cinco emoções das 28 crianças com desenvolvimento típico, usando LDA com base na matriz de covariância completa . A Figura 8 e as Tabelas 3 e 4 mostram que os métodos PCA e FNCA são intercambiáveis para redução de dimensionalidade, alcançando ACC médio de 85,29% e 85,75%, respectivamente.

Tabela 3. Desempenho do sistema de reconhecimento de cinco emoções usando FNCA mais LDA com base em matrizes de covariância.

	Nojo	Temer	Felicidade	Tristeza	Surpresa
TPR (%)	90,02 ± 1,91	87,25 ± 2,37	87,55 ± 2,12	74,69 ± 2,77	86,93 ± 2,12
FPR (%)	3,07 ± 0,49	4,17 ± 0,57	4,61 ± 0,69	3,42 ± 0,78	2,93 ± 0,50
ACC (%)			85,29 ± 1,16		
Capa (%)			81,26 ± 1,46		

ACC, precisão; TPR, taxa de verdadeiro positivo; FPR, taxa de falsos positivos.

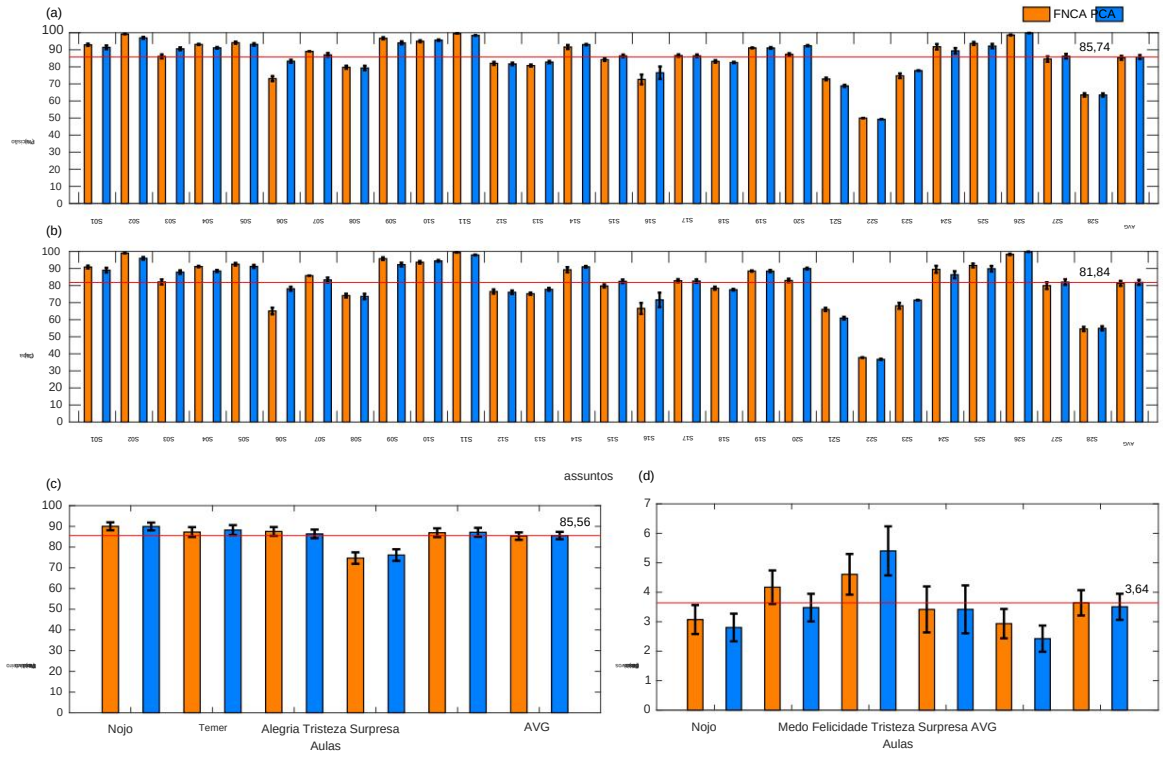

Figura 8. Desempenho da classificação de cinco emoções aplicando análise de componentes principais (PCA) e Análise de Componentes de Vizinhança Rápida (FNCA) para redução de dimensionalidade.

Tabela 4. Desempenho do sistema proposto para reconhecimento de cinco emoções utilizando PCA com 60 componentes mais LDA baseado em matrizes de covariância completas.

	Nojo	Temer	Felicidade	Tristeza	Surpresa
TPR (%)	89,93 ± 1,88	88,22 ± 2,38	86,36 ± 2,09	76,15 ± 2,80	87,14 ± 2,15
FPR (%)	2,80 ± 0,47	3,48 ± 0,47	5,40 ± 0,83	3,42 ± 0,81	2,42 ± 0,44
ACC (%)	85,75 ± 1,16				
Capa (%)	81,84 ± 1,46				

ACC, precisão; TPR, taxa de verdadeiro positivo; FPR, taxa de falsos positivos.

Vale ressaltar que o PCA melhorou o desempenho do sujeito S06 (ACC de 83,27% e Kappa de 78,10% para PCA, e ACC de 73,11% e Kappa de 65,09% para FNCA), conforme mostrado na Figura 8a,b. Essa melhoria pode ser consequência da robustez do PCA para extração de recursos sob condições controladas de pose e imagem [9]. Além disso, observe que para o sujeito S22 o menor desempenho foi obtido pela aplicação de PCA (ACC de 49,27% e Kappa de 36,75%) e FNCA (ACC de 49,98% e Kappa de 37,78%). No entanto, valores de ACC superiores a 85,74% foram alcançados para um total de 14 crianças pela aplicação de PCA ou FNCA, sendo o maior ACC de 99,78% para o sujeito S26 usando PCA. Rotular vetores de características de padrões de emoção é um desafio, principalmente quando os rótulos são atribuídos ao longo de períodos de tempo, pois podem existir alguns atrasos entre a percepção e a atribuição manual de rótulos. Da mesma forma, um estímulo visual pode produzir diferentes emoções faciais, assim, o procedimento manual rotulado é subjetivo. Então, um método automático para rotulagem de conjunto de recursos pode ser adequado para obter rótulos confiáveis e, portanto, um modelo de classificação para LDA.

Por outro lado, a Figura 8c mostra que quatro emoções (desgosto, medo, alegria e surpresa) foram reconhecidas com sucesso pela aplicação de PCA e FNCA, alcançando TPR > 85%, enquanto a tristeza foi classificada com menor sensibilidade (TPR de 74,69% e 76,15% para FNCA e PCA, respectivamente) pelo sistema proposto. Como destaque, baixos valores de FPR (ÿ5,40%) foram obtidos para reconhecer todos os

emoções. Em geral, os resultados alcançados mostram que PCA e FNCA são intercambiáveis para obter as características que aumentam a discriminação de emoções, descobrindo aquelas regiões que são informativas em termos de expressões [9]. Semelhante a outros estudos [52], a emoção tristeza foi menos reconhecida (ACC < 77%) do que as demais emoções, pois cada criança expressou tristeza de uma maneira diferente, produzindo uma grande variabilidade intraclasse para tristeza.

Muitos pesquisadores [1] realizaram estudos para inferir emoções faciais sobre imagens visuais usando apenas três das seis expressões básicas, especificamente felicidade, surpresa e tristeza, pois essas emoções são mais fáceis de serem reconhecidas por humanos e têm sido usadas com sucesso para gerar expressões coesas entre os participantes. Da mesma forma, outros estudos usando imagens visuais relatam que as taxas de reconhecimento de expressões como medo ou nojo podem ser muito baixas, na faixa de 40 a 50%. Como destaque, cinco emoções básicas foram reconhecidas pelo nosso sistema proposto baseado em câmera infravermelha, alcançando o maior desempenho tanto para desgosto (TPR de 89,93%) quanto para medo (TPR de 88,22%), seguido de felicidade (TPR de 86,36%) e surpresa (TPR de 87,14%), como mostrado na Figura 8c e Tabela 4. Em concordância com outros estudos [17,21], nossos resultados mostram que as imagens térmicas são promissoras para análise e reconhecimento de emoções faciais.

Da mesma forma, este banco de dados foi usado como um conjunto de treinamento do nosso sistema de reconhecimento para inferir emoções em 17 crianças, enquanto inicialmente permaneciam na frente do N-MARIA por um período de tempo, e depois que ele foi descoberto. As Figuras 9a–d e 10a,b mostram emoções inferidas pelo nosso sistema para as dezessete crianças durante os experimentos Baseline e Test, sendo felicidade e surpresa as saídas mais frequentes. Emoções como nojo, medo e tristeza foram poucas inferidas pelo sistema, concordando com o que as crianças relataram, como mostra a Figura 10c.

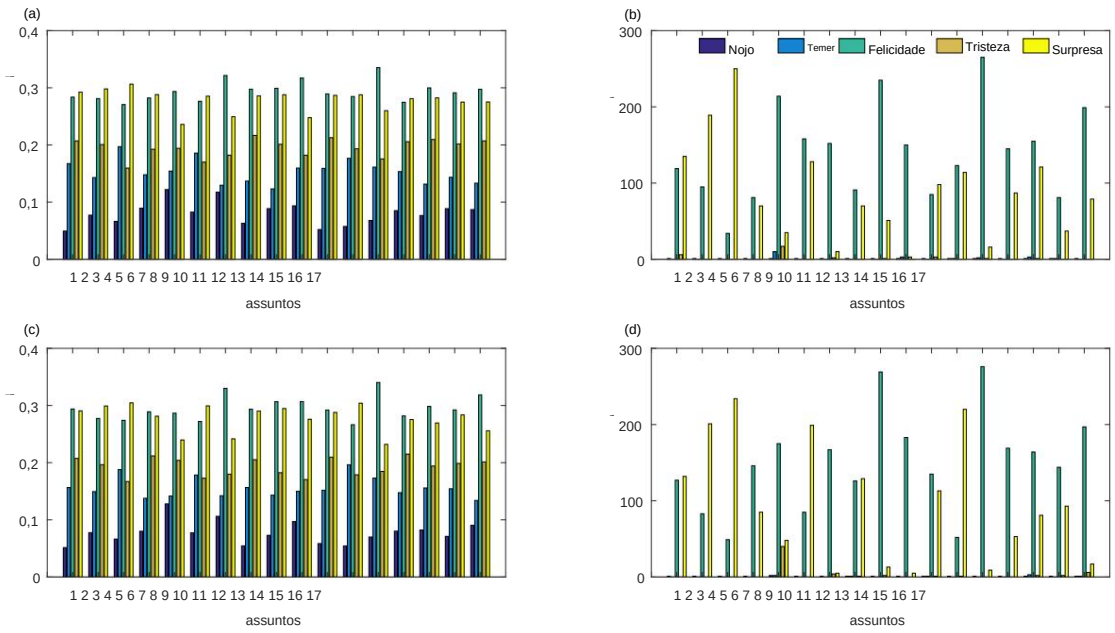

Figura 9. Reconhecimento automático de emoções sobre padrões desconhecidos. (a,b) são emoções inferidas obtidas durante a fase de linha de base do experimento; (c,d) são decisões emocionais alcançadas para a fase de teste do experimento.

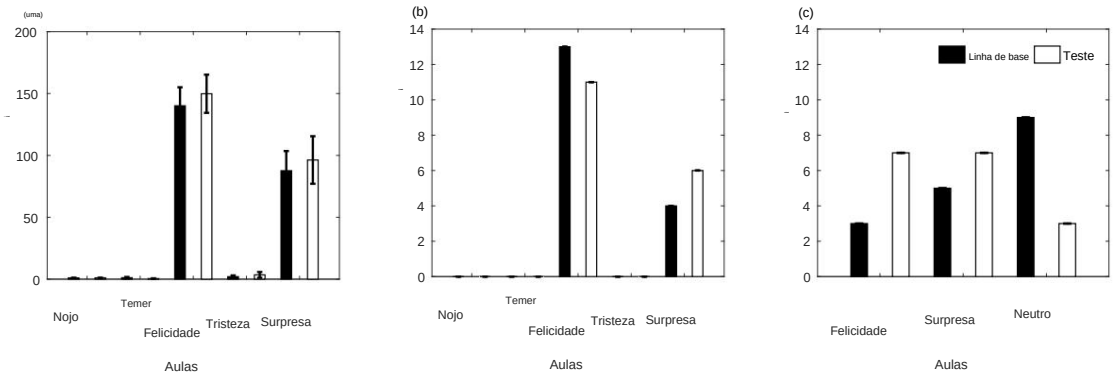

Figura 10. Resumo das emoções das crianças antes e depois de verem o robô pela primeira vez. **(a)** e **(b)** são emoções inferidas pelo sistema de reconhecimento; **(c)** emoções relatadas pelas crianças.

A Figura 10c mostra que uma emoção neutra foi relatada por mais crianças para N-MARIA cobertas, enquanto elas sentiram surpresa e felicidade ao ver N-MARIA pela primeira vez. Em contrapartida, felicidade e surpresa foram um pouco mais inferidas após a descoberta do robô para as crianças, o que também concordou com o que as crianças relataram. Vale ressaltar que as emoções podem variar entre as crianças para um mesmo estímulo visual, como um robô social. Por esta razão, um único indicador de avaliação não é suficiente para interpretar com precisão as respostas das crianças a um robô enquanto estão interagindo. Em seguida, um questionário foi respondido por cada criança, relatando suas emoções durante a Linha de Base e o Teste (ver Seção 3.1). No entanto, as autoavaliações têm problemas de validade e corroboração, conforme descrito na Ref. [1], pois os participantes podem relatar de forma diferente de como estão realmente pensando ou sentindo. Então, não é trivial atribuir as respostas das crianças aos seus verdadeiros comportamentos. Da mesma forma, a emoção produzida por cada criança pode ser condicionada, pois ela sabe que está sendo observada pelo robô, pelo pai e pelo especialista que conduz o experimento. Observe também que um banco de dados existente de 28 crianças com desenvolvimento típico (idade: 7-11 anos) foi usado em nosso estudo para treinar nosso sistema proposto para inferir uma das cinco emoções básicas (nojo, medo, felicidade, tristeza, surpresa) produzidas por 17 crianças durante a interação com N-MARIA. No entanto, o processamento da imagem visual e térmica pode ser afetado devido à qualidade da imagem de entrada, pois depende das condições de iluminação e da distância entre criança e robô durante a interação.

5. Discussão

Vários estudos foram realizados para reconhecer algumas das seis emoções aceitas pela teoria psicológica: surpresa, medo, nojo, raiva, felicidade e tristeza [10,53-55]. Por exemplo, o movimento facial e o tom da fala têm mostrado em sistemas de reconhecimento seu papel relevante para inferir essas emoções supracitadas, alcançando precisão (ACC) de 80% a 98% [56–58] e de 70% a 75% [53 ,59,60], respectivamente. Embora as expressões faciais forneçam pistas importantes sobre as emoções, é necessário medir, por fluxo óptico, os movimentos de músculos faciais específicos por meio de marcadores localizados na face [53,57,58]. No entanto, essa técnica sem contato pode não ser confortável para estudar ou inferir emoções de crianças com TEA, pois essas crianças apresentam alta sensibilidade cutân

Seguindo essa abordagem de técnicas sem contato, existem algumas APIs (Application Program Interface) que permitem a detecção facial, como a “Emotion API” da Microsoft ou a Oxford API [61,62]. Alguns trabalhos, alternativamente, utilizam outras técnicas de processamento de imagem, como encontrar a região do termograma dentro de uma faixa de temperatura para detectar a face [18]. Outra abordagem é manter a face em posição fixa utilizando um suporte para queixo ou dispositivo de apoio de cabeça [41,63]. Além disso, na Ref. [64] os autores apresentam uma solução utilizando métodos de classificação de formas supervisionadas e redes neurais aplicados ao termograma facial.

Os sistemas de reconhecimento de emoções baseados em IRTI têm, de fato, mostrado resultados promissores. A Tabela 5 mostra alguns estudos que usam substituição de ROI e análise de emoções adicionais, embora usando outras técnicas

em comparação com a nossa proposta. Infelizmente, os resultados apresentados são um tanto dispersos, e não é possível fazer uma comparação justa entre eles, devido às diferentes imagens utilizadas no estudos. Na Ref. [18] os autores propuseram técnicas para selecionar ROIs faciais e classificar emoções usando uma câmera FLIR A310. Para detectar o rosto, as temperaturas do termograma entre 32 ÿC a 36 ÿC foram usados para definir a posição da face. As posições de ROIs foram ainda calculadas por proporções com base na largura da cabeça. Todos os outros pontos de temperatura foram considerados de fundo. Adicionalmente, para classificação de emoções, o sistema foi calibrado usando uma linha de base (estado neutro) que compensa a emoção induzida pela aplicação do algoritmo fuzzy e, assim, calibrar a imagem da emoção induzida. Usando a linha de base, a temperatura é inferida pelas regras IF-THEN para calibrar as imagens térmicas para as seguintes emoções induzidas: alegria, desgosto, raiva, medo e tristeza. Em seguida, uma hierarquia de cima para baixo classificador foi usado para analisar a classificação da emoção, atingindo 89,9% de taxa de sucesso.

Tabela 5. Comparação entre algumas estratégias utilizadas para inferir emoções usando Infrared Thermal Imagiologia (IRTI).

Estudos	Voluntários	Idade	ACC (%)	Resumo
Cruz-Albaran et ai. [18]	25	19 a 33	89,90	Algoritmo difuso, regras IF-THEN e top-down classificador hierárquico. As emoções analisadas foram alegria, desgosto, raiva, medo e tristeza.
Basu et ai. [13]	26	11 a 32	87,5	Extração de recursos de histograma e suporte multiclasse máquina vetorial (SVM). O Facial Térmico Kotani O banco de dados de expressões foi usado para detectar emoções de raiva, medo, alegria e tristeza.
Nhan e Chau [41]	12	21 a 27	80,0	Comparação entre os estados basal e afetivo. Alto e baixa excitação e valência são comparados com o linha de base.
Wang et ai. [65]	38	17 a 31	62,90	Deep Boltzmann Machine para encontrar positivo ou negativo valência.
Bijalwan et ai. [66]	1	N / D	97,0	Modelo para reconhecimento de expressão em imagens térmicas. Aplicação do PCA para reconhecer felicidade, raiva, desgosto, tristeza e emoções neutras.
Yoshitomi et ai. [67]	1	N / D	90,0	Redes neurais e algoritmos de retropropagação que reconhecer emoções de felicidade, surpresa e neutralidade Estado.
Kosonogov et ai. [63]	24	20 a 24	N/A	Estudou a variação térmica da ponta do nariz em voluntários com imagens do International Affective Picture System (IAPS) que descobriram que imagens positivas ou negativas mostravam mais mudança de temperatura em comparação com imagens neutras
Vukadinovic e Pânico [68]	200	18 a 50	N/A	Algoritmo para encontrar ROIs faciais. A Viola-Jones adaptado algoritmo (que aplica GentleBoost em vez de AdaBoost) foi usado. Para extração de características faciais Gabor wavelet filtro foi usado. Não foram estudadas aulas de emoção, apenas os pontos faciais para detecção de ROI.
Bharatharaj et al. [62]	9	6 a 16	N/A	AMRM (método de ensino indireto) foi estudado usando um robô inspirado em papagaios e a API de emoção Oxford para reconhecer e classificar emoções em crianças com TEA. O máximo de eles pareciam estar felizes com o robô.
Mehta et ai. [61]	3	N / D	93,8%	O sistema Microsoft HoloLens (MHL) foi usado para detectar emoções alcançando uma alta precisão para detectar a felicidade, tristeza, raiva, surpresa e emoções neutras.
Nossa proposta	28	7 a 11	85,75%	PCA e LDA foram usados em nosso banco de dados, publicado em [17], reconhecer alegria, tristeza, medo, surpresa e desgosto.

ACC, precisão; N/A significa que a idade ou ACC não foram informados.

A imagem térmica infravermelha funcional (fITI) foi utilizada nas Refs. [21,63], que é considerado ser uma técnica promissora para inferir emoções por meio de respostas autonômicas. Da mesma forma, outro estudo foi realizado usando fITI para comparar classificações subjetivas de imagens exibidas para os voluntários [63], onde essas imagens foram categorizadas em desagradáveis, neutras e agradáveis. Então, enquanto os voluntários observando essas fotos, os autores coletaram a temperatura da ponta do nariz (havia um apoio de queixo para manter o rosto corretamente localizado na imagem da câmera), que é um dos lugares mais prováveis de mudar temperatura quando a pessoa está sob algum tipo de emoção [17]. Como resultado, eles descobriram que as imagens

que evocam emoções (não importa se é uma emoção positiva ou negativa) foram mais suscetíveis a produzir variação térmica, enquanto a diferença para as imagens neutras não foi tão grande quanto as demais. Assim, suas descobertas demonstram que o fITI pode ser uma ferramenta útil para inferir emoções em humanos.

Outra pesquisa interessante [68] localiza pontos faciais em imagens visuais em tons de cinza usando classificadores impulsionados baseados em recursos Gabor, nos quais os autores usaram uma versão adaptada do algoritmo Viola-Jones, usando GentleBoost em vez de AdaBoost, para detectar a face. Além disso, a wavelet Gabor foi usada para extração de características, detectando 20 ROIs que representam os pontos de características faciais. Toda essa detecção foi feita de forma automática e sem contato usando a detecção de íris e boca. Essas duas partes foram detectadas dividindo o rosto em duas regiões e calculando proporções para encontrar essas regiões (íris e boca). A partir disso, todos os outros ROIs foram calculados usando proporções. Sua taxa de sucesso foi alta, pois o algoritmo alcançou 93% de taxa de sucesso usando o banco de dados Cohn-Kanade, que tem fotos inexpressivas de 200 pessoas. Embora a transformada wavelet de Gabor seja um método representativo para extrair feições locais, leva muito tempo e tem uma grande dimensão de feições.

Outro método foi proposto na Ref. [65], onde uma máquina de Boltzmann profunda (DBM) foi aplicada para reconhecer emoções a partir de imagens térmicas faciais, utilizando um banco de dados prévio e com a participação de 38 voluntários adultos. Sua avaliação consistiu em encontrar a valência da emoção, que poderia ser positiva ou negativa, e sua taxa de acerto chegou a 62,9%. Em seu estudo, como a face e o fundo têm temperaturas diferentes, eles foram divididos aplicando o algoritmo de limiar Otsu para binarizar as imagens. Em seguida, as curvas de projeção (vertical e horizontal) foram calculadas para encontrar o maior gradiente e detectar o limite da face.

Adicionalmente, um modelo de reconhecimento de expressão usando imagens térmicas de um voluntário adulto foi aplicado na Ref. [66]. Esses autores utilizaram eigenfaces para extração de características das imagens faciais do voluntário por meio de PCA para reconhecer cinco emoções (alegria, raiva, desgosto, triste e neutra). Como destaque, essa proposta atingiu uma acurácia próxima a 97%, no qual trabalharam, aplicaram autovalores e eigenfaces, treinaram o sistema com um conjunto de imagens, utilizaram PCA para reduzir a dimensionalidade e classificador de distância para reconhecer a emoção.

Na Ref. [13], os autores conseguiram atingir 81,95% de precisão usando extração de características de histograma combinada com SVM multiclasse sobre imagens térmicas de 22 voluntários no banco de dados Kotani Thermal Facial Expression (KTFE), e quatro classes foram estudadas: felicidade, tristeza, medo e raiva. Eles usaram técnicas de pré-processamento para preparar a imagem para aplicar Viola-Jones e para melhorar ainda mais a imagem, eliminando o ruído da imagem e usando a equalização adaptativa de histograma de contraste limitado. Para detecção de ROI, foi usada uma segmentação baseada em razão.

Além disso, o reconhecimento dos estados de linha de base e afetivos foi realizado na Ref. [41], onde, para detectar a face, utilizaram um apoio de cabeça para mantê-la na posição correta, além de um ponto de referência (localizado no topo da cabeça), que estava cerca de 10 ÿC mais frio que a temperatura da pele . Para encontrar as ROIs, utilizou-se o ponto de referência e aplicou-se um limiar radiométrico. Em caso de perda do ponto de referência, este foi corrigido manualmente pelos pesquisadores.

Outro estudo [67] aplicou IRTI em uma voluntária adulta do sexo feminino, e os algoritmos de Redes Neurais e Backpropagation foram usados para reconhecer emoções, como felicidade, surpresa e estado neutro, atingindo um ACC de 90%. Para encontrar a face eles usaram segmentação Otsu, e o diâmetro do Feret foi encontrado na imagem binária junto com o centro de gravidade da imagem binária. Em seguida, após a segmentação da imagem, as posições do rosto baseadas em FACS-AU foram utilizadas para determinar a variação do calor e, assim, a emoção.

Outro trabalho [61] mostra um estudo sobre várias abordagens de reconhecimento de emoções e detecção facial, como aprendizado de máquina e processo baseado em recursos geométricos, além de SVM e uma diversidade de outros classificadores. Eles também apresentam o uso do Microsoft HoloLens (MHL) para detectar emoções humanas usando um aplicativo que foi desenvolvido para usar o MHL para detectar rostos e reconhecer emoções de pessoas que o enfrentam. O conjunto de emoções que trabalharam foi composto por alegria, tristeza, raiva, surpresa e neutro. Além disso, eles usaram uma webcam para detectar emoções e comparar com o resultado usando MHL. O sistema com MHL pode alcançar resultados muito melhores do que os trabalhos anteriores

e teve uma precisão notável provavelmente devido aos sensores acoplados ao HoloLens, atingindo uma precisão de 93,8% em MHL, usando a “API Emotion” da Microsoft.

Na Ref. [62], os autores usaram um robô inspirado em papagaios (KiliRo) para interagir com crianças com TEA simulando um conjunto de comportamentos autônomos. Eles testaram o robô por cinco dias consecutivos em uma clínica. As expressões das crianças ao interagir com o robô foram analisadas pela API de emoções da Oxford, permitindo que elas fizessem um sistema automatizado de detecção facial, reconhecimento de emoções e classificação.

Algumas obras, como a Ref. [69], mostram o uso do aprendizado profundo para detectar a face da criança e inferir a atenção visual em um robô durante a terapia de CRI. Os autores utilizaram o robô NAO da Softbank Robotics, que possui duas câmeras de baixa resolução que foram utilizadas para tirar fotos e gravar vídeos. Eles também usaram o sistema de detecção e rastreamento de rosto embutido no NAO para os experimentos clínicos. Um total de 6 crianças participaram do experimento, no qual imitaram alguns movimentos de robôs. As crianças tiveram 14 encontros ao longo de um mês, e os experimentos propriamente ditos começaram 7 dias após o encontro preliminar, a fim de evitar o efeito de novidade nos resultados. Diferentes técnicas e classificadores de deep learning foram utilizados, podendo atingir uma taxa média de atenção das crianças de 59,2%.

Abordagens baseadas em aprendizado profundo têm se mostrado promissoras para reconhecimento de emoções, determinando características e classificadores sem supervisores especializados [10]. No entanto, abordagens convencionais ainda estão sendo estudadas para uso em sistemas embarcados de tempo real devido à sua baixa complexidade computacional e alto grau de precisão [70], embora para esses sistemas os métodos de extração e classificação de características devam ser projetados pelo programador e não podem ser otimizado para aumentar o desempenho [71]. Além disso, vale a pena mencionar que as abordagens convencionais exigem poder computacional e memória relativamente menores do que as abordagens baseadas em aprendizado profundo [10]. Da mesma forma, os recursos Gabor são muito populares para classificação de expressões faciais e reconhecimento facial, devido ao seu alto poder discriminativo [72,73], mas a complexidade computacional e o requisito de memória os tornam menos adequados para implementação em tempo real.

Nosso sistema é composto por hardware de baixo custo e métodos de baixo custo computacional para processamento de imagens visuais e térmicas, e reconhece cinco emoções, alcançando 85,75% de precisão. Para o nosso sistema, propusemos um método baseado no erro de probabilidade para localizar com precisão pontos de referência específicos do assunto, levando em consideração os critérios de especialistas treinados. Como destaque, nossa proposta pode encontrar quadro a quadro o ROI facial mais bem localizado usando o algoritmo Viola-Jones e ajustar a localização de seus ROIs faciais circundantes. Como outra novidade, nossa proposta baseada em erro de probabilidade mostrou robustez e boa precisão para localizar ROIs faciais em imagens térmicas, que foram coletadas enquanto crianças com desenvolvimento típico interagiam com um robô social. Como outros achados, estendemos um banco de dados existente de cinco emoções faciais a partir de imagens térmicas, para inferir emoções desconhecidas geradas enquanto as crianças interagiam com o robô social, usando nosso sistema de reconhecimento baseado em PCA e LDA, alcançando assim resultados que concordavam com o escrito relatos de crianças.

Como limitação, nosso sistema não é capaz de rastrear os movimentos da cabeça, adicionando assim um método de rastreamento facial, como feito pela Ref. [74], pode tornar robusta nossa proposta de referências faciais em cenários não controlados, como aplicativos móveis para interação de crianças e robôs sociais. Geralmente, os conjuntos de dados de emoções faciais com seis emoções básicas contêm apenas participantes adultos, mas há muito poucos bancos de dados coletados em crianças com desenvolvimento típico (com idades entre 7 e 11 anos) por meio de câmera infravermelha, contendo as emoções básicas. Então, é um desafio usar uma grande quantidade de exemplos durante a fase de treinamento de um sistema de reconhecimento para inferir emoções de crianças de 7 a 11 anos enquanto elas interagem com um robô, por exemplo. Além disso, mais testes com maior número de voluntários devem ser realizados, incluindo crianças com TEA.

6. conclusões

Um sistema computacional de baixo custo para reconhecimento de emoções infantis de forma discreta foi proposto neste estudo, que é composto por câmeras de baixo custo, possibilitando sua extensão para pesquisas em países em desenvolvimento. Numa primeira fase, a nossa proposta foi testada em termos visuais e térmicos.

imagens de crianças interagindo com o robô social móvel N-MARIA, alcançando resultados promissores (85,75% de acerto) para localizar pontos de referência faciais específicos, bem como reconhecer (ou inferir) cinco emoções. Todas as crianças tiveram uma interação esperançosa com o robô, o que demonstrou que nosso sistema é útil para estimular emoções positivas nas crianças e capaz de desencadear uma interação proveitosa com elas. Em trabalhos futuros, esta proposta será integrada ao N-MARIA, visando conhecer online a emoção das crianças e tomar decisões de controle baseadas nas emoções. Além disso, outros métodos serão explorados para rastreamento facial, a fim de reduzir a influência da postura da cabeça durante o reconhecimento de emoções. Além disso, métodos não supervisionados para atribuição automática de rótulos e aprendizado de classificadores serão avaliados em nosso conjunto de dados para obter um sistema de reconhecimento robusto para processar padrões de alta incerteza, como expressões faciais e emoções.

Contribuições dos Autores: Conceituação, CG, CV, DD-R. e DF; Metodologia, GC e DF; Software, DF, VB, AF, GB, CV e DD-R.; Análise Formal, GC e DD-R.; Redação—Preparação do Projeto Original, CG, DD-R., CV e DF; Redação—Revisão e Edição, EC e TB-F.

Financiamento: Esta pesquisa foi financiada pela FAPES/Brasil, números 72982608 e 645/2016.

Agradecimentos: Os autores agradecem o apoio financeiro da CAPES, CNPq e FAPES/Brasil (números dos projetos: 72982608 e 645/2016) e UFES pelo apoio técnico.

Conflitos de interesse: Os autores declaram não haver conflito de interesse. Os financiadores não tiveram nenhum papel no desenho do estudo; na coleta, análise ou interpretação dos dados; na redação do manuscrito, ou na decisão de publicar os resultados.

Referências

1. Gunes, H.; Celiktutan, O.; Sariyanidi, E. Live demonstrações públicas interativas humano-robô com emoção automática e previsão de personalidade. Philos. Trans. R. Soc. B **2019,** 374, 20180026. [CrossRef] [PubMed]
2. Kim, ES; Berkovits, LD; Bernier, EP; Leyzberg, D.; Shic, F.; Paulo, R.; Scassellati, B. Robôs Sociais como Reforçadores Incorporados do Comportamento Social em Crianças com Autismo. J. Autismo Dev. Desordem. **2012,** 43, 1038-1049. [CrossRef] [PubMed]
3. Valadao, C.; Caldeira, E.; Bastos-Filho, T.; Frizera-Neto, A.; Carelli, R. Um novo controlador para um Smart Walker Baseado na Formação Humano-Robô. Sensores **2016,** 16, 1116. [CrossRef] [PubMed]
4. Picard, R.; Vyzas, E.; Healey, J. Rumo à inteligência emocional da máquina: análise do estado fisiológico afetivo . Trans. IEEE Padrão Anal. Mach. Intel. **2001,** 23, 1175-1191. [CrossRef]
5. Conn, K.; Liu, C.; Sarkar, N.; Stone, W.; Warren, Z. Tecnologias de intervenção assistida sensíveis ao afeto para crianças com autismo: Uma abordagem individual específica. In Proceedings of the RO-MAN 2008—The 17th IEEE International Symposium on Robot and Human Interactive Communication, Munique, Alemanha, 1–3 de agosto de 2008.
6. Shier, WA; Yanushkevich, SN Biometria na interação homem-máquina. In Proceedings of the 2015 International Conference on Information and Digital Technologies, Zilina, Eslováquia, 7–9 de julho de 2015. doi:10.1109/dt.2015.7222989.
7. Goulart, C.; Valadao, C.; Caldeira, E.; Bastos, T. Avaliação do sinal cerebral de crianças com Transtorno do Espectro Autista na interação com um robô social. Biotecnologia. Res. Inovação **2018.** [CrossRef]
8. Latif, MT; Yusof, M.; Fatai, S. Detecção de Emoção de Impressão Facial Térmica com base em Recursos GLCM. ARPN J. Eng. Aplic. Sci. **2016,** 11, 345-349.
9. Sariyanidi, E.; Gunes, H.; Cavallaro, A. Análise automática do afeto facial: Uma pesquisa de registro, representação e reconhecimento. Trans. IEEE Padrão Anal. Mach. Intel. **2014,** 37, 1113-1133. [CrossRef] [PubMed]
10. Ko, B. Uma breve revisão do reconhecimento de emoções faciais com base em informações visuais. Sensores **2018,** 18, 401. [CrossRef]
11. Rusli, N.; Sidek, SN; Yusof, HM; Latif, MHA Avaliação não invasiva de estados afetivos em indivíduos com Transtorno do Espectro Autista: Uma Revisão. Nos Anais do IFMBE; Springer: Cingapura, 2015; págs. 226-230.
12. Petrantonakis, PC; Hadjileontiadis, LJ Reconhecimento de emoções de EEG usando cruzamentos de ordem superior. Trans. IEEE Inf. Tecnol. Biomédico. **2009,** 14, 186-197. [CrossRef]

13. Basu, A.; Routray, A.; Merda, S.; Deb, AK Reconhecimento de emoção humana a partir de imagem térmica facial com base em recurso estatístico fundido e SVM multiclasse. In Proceedings of the 2015 Annual IEEE India Conference (INDICON), Nova Delhi, Índia, 17 a 20 de dezembro de 2015.
14. Ghimire, D.; Jeong, S.; Lee, J.; Park, SH Reconhecimento de expressão facial com base em características específicas da região local e máquinas de vetor de suporte. Multimed. Ferramentas Aplic. **2017,** 76, 7803-7821. [CrossRef]
15. Perikos, I.; Paraskevas, M.; Hatzilygeroudis, I. Reconhecimento de Expressão Facial Usando Sistemas Adaptativos de Inferência Neuro-fuzzy. In Proceedings of the 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), Cingapura, 6 a 8 de junho de 2018; págs. 1–6. [CrossRef]
16. Feliz, S.; Routray, A. Reconhecimento automático de expressão facial usando recursos de manchas faciais salientes. Trans. IEEE Afeta. Computar. **2014,** 6, 1-12. [CrossRef]
17. Goulart, C.; Valadao, C.; Delisle-Rodriguez, D.; Caldeira, E.; Bastos, T. Análise das emoções em crianças através emissividade facial de imagens térmicas infravermelhas. PLoS ONE **2019,** 14, e0212928.. [CrossRef] [PubMed]
18. Cruz-Albarn, IA; Benitez-Rangel, JP; Osornio-Rios, RA; Morales-Hernandez, LA Detecção de emoções humanas com base em um sistema térmico inteligente de imagens termográficas. Física infravermelha. Tecnol. **2017,** 81, 250-261. [CrossRef]
19. Wang, S.; Shen, P.; Liu, Z. Reconhecimento de expressão facial de imagens térmicas infravermelhas usando diferença de temperatura por votação. In Proceedings of the 2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems, Hangzhou, China, 30 de outubro a 1º de novembro de 2012; Volume 1, pp. 94–98.
20. Pop, FM; Gordan, M.; Florea, C.; Vlaicu, A. Abordagem baseada em fusão para reconhecimento facial térmico e visível sob variação de pose e expressividade. In Proceedings of the 9th RoEduNet IEEE International Conference, Sibiu, Romênia, 24–26 de junho de 2010; págs. 61-66.
21. Ioannou, S.; Gallese, V.; Merla, A. Imagens infravermelhas térmicas em psicofisiologia: potencialidades e limites. Psicofisiologia **2014,** 51, 951-963. [CrossRef] [PubMed]
22. Zheng, Y. Detecção de rosto e detecção de óculos para reconhecimento térmico de rosto. SPIE Proc. **2012,** 8300, 83000C.
23. Wang, S.; Liu, Z.; Eu contra.; Lv, Y.; Wu, G.; Peng, P.; Chen, F.; Wang, X. Um Banco de Dados de Expressão Facial Visível e Infravermelho Natural para Reconhecimento de Expressão e Inferência de Emoções. Trans. IEEE Multimed. **2010,** 12, 682-691. [CrossRef]
24. Choi, JS; Bang, J.; Heo, H.; Park, K. Avaliação do medo usando medição não intrusiva de multimodal sensores. Sensores **2015,** 15, 17507-17533. [CrossRef] [PubMed]
25. Lajevardi, SM; Hussain, ZM Reconhecimento automático de expressões faciais: Extração e seleção de recursos. Processo de vídeo de imagem de sinal. **2012,** 6, 159-169. [CrossRef]
26. Jabid, T.; Kabir, MH; Chae, O. Reconhecimento de expressão facial robusto baseado em padrão direcional local. ETRI J. **2010,** 32, 784-794. [CrossRef]
27. Kabir, MH; Jabid, T.; Chae, O. Um descritor de face baseado em variação de padrão direcional local (LDPv) para reconhecimento de expressão facial humana. In Proceedings of the 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, Boston, MA, EUA, 29 de agosto a 1 de setembro de 2010; págs. 526-532.
28. Shan, C.; Gong, S.; McOwan, PW Reconhecimento de expressão facial robusto usando padrões binários locais. In Proceedings of the IEEE International Conference on Image Processing 2005, Genova, Itália, 14 de setembro de 2005; Volume 2, pp. II-370.
29. Shan, C.; Gritti, T. Aprendizagem discriminativa LBP-Histograma Bins para reconhecimento de expressão facial. In Proceedings of the British Machine Vision Conference 2008, Leeds, Reino Unido, 1–4 de setembro de 2008; págs. 1–10.
30. Song, M.; Tao, D.; Liu, Z.; Li, X.; Zhou, M. Recursos de proporção de imagem para aplicação de reconhecimento de expressão facial. Trans. IEEE Sistema Homem Cibernético. Parte B Cibern. **2009,** 40, 779-788. [CrossRef]
31. Zhang, L.; Tjondronegoro, D. Reconhecimento de expressão facial usando recursos de movimento facial. Trans. IEEE Afeta. Computar. **2011,** 2, 219-229. [CrossRef]
32. Viola, P.; Jones, MJ Detecção de rosto em tempo real robusta. Int. J. Computação. Vis. **2004,** 57, 137-154. [CrossRef]
33. Jiang, B.; Martinez, B.; Valstar, MF; Pantic, M. Fusão em nível de decisão de regiões específicas de domínio para reconhecimento de ação facial. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Estocolmo, Suécia, 24–28 de agosto de 2014; págs. 1776-1781.
34. Shan, C.; Gong, S.; McOwan, PW Reconhecimento de expressão facial baseado em padrões binários locais: Um estudo abrangente. Imagem Vis. Computar. **2009,** 27, 803-816. [CrossRef]

35. Kazemi, V.; Sullivan, J. Um alinhamento de face de um milissegundo com um conjunto de árvores de regressão. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, EUA, 24–27 de junho de 2014; págs. 1867-1874.
36. Zhao, X.; Zhang, S. Reconhecimento de expressão facial baseado em padrões binários locais e discriminante de kernel isomapa. Sensores **2011,** 11, 9573-9588. [CrossRef] [PubMed]
37. Yang, J.; Wang, X.; Han, S.; Wang, J.; Parque, DS; Wang, Y. Reconhecimento de expressão facial em tempo real aprimorado com base em uma nova codificação de gradiente local equilibrada e simétrica. Sensores **2019,** 19, 1899. [CrossRef] [PubMed]
38. Giacinto, AD; Brunetti, M.; Sepede, G.; Ferretti, A.; Merla, A. Assinatura térmica do condicionamento do medo no transtorno de estresse pós-traumático leve. Neurociência **2014,** 266, 216-223. [CrossRef] [PubMed]
39. Marzec, M.; Koprowski, R.; Wróbel, Z. Métodos de localização facial em termogramas. Biocibern. Biomédico. Eng. **2015,** 35, 138-146. [CrossRef]
40. Trujillo, L.; Olague, G.; Hammoud, R.; Hernandez, B. Localização automática de recursos em imagens térmicas para reconhecimento de expressões faciais. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)—Workshops, San Diego, CA, EUA, 21–23 de setembro de 2005. doi:10.1109/cvpr.2005.415.
41. Nhan, B.; Chau, T. Classificando Estados Afetivos Usando Imagens Infravermelhas Térmicas da Face Humana. Trans. IEEE Biomédico. Eng. **2010,** 57, 979-987. [CrossRef]
42. Bradski, G. A biblioteca OpenCV. J. Softw do Dr. Dobb. Ferramentas **2000,** 25, 120–125.
43. Malis, E.; Vargas, M. Compreensão mais profunda da decomposição de homografia para controle baseado em visão. Ph.D. Tese, INRIA, Sophia Antipolis Cedex, França, 2007.
44. Budzier, H.; Gerlach, G. Calibração de câmeras infravermelhas térmicas não refrigeradas. J. Sens. Sens. Syst. **2015,** 4, 187-197. [CrossRef]
45. Martínez, AM; Kak, AC Pca versus lda. Trans. IEEE Padrão Anal. Mach. Intel. **2001,** 23, 228-233. [CrossRef]
46. Friedman, JH Análise discriminante regularizada. Geléia. Estado. Associação **1989,** 84, 165-175. [CrossRef]
47. Kwon, OW; Chan, K.; Hao, J.; Lee, TW Reconhecimento de emoções por sinais de fala. In Proceedings of the Eighth European Conference on Speech Communication and Technology, Genebra, Suíça, 1-4 de setembro de 2003.
48. Bamidis, PD; Frantzidis, CA; Konstantinidis, EI; Luneski, A.; Lithari, C.; Klados, MA; Bratsas, C.; Papadelis, CL; Pappas, C. Uma abordagem integrada ao reconhecimento de emoções para inteligência emocional avançada. Na Conferência Internacional sobre Interação Humano-Computador; Springer: Berlim/Heidelberg, Alemanha, 2009, pp. 565–574.
49. Ververidis, D.; Kotropoulos, C.; Pitas, I. Classificação automática da fala emocional. In Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, QC, Canadá, 17–21 de maio de 2004; Volume 1, pp. I-593.
50. Hsu, CW; Lin, CJ Uma comparação de métodos para máquinas vetoriais de suporte multiclasse. Trans. IEEE Neural Rede **2002,** 13, 415-425. [PubMed]
51. Japkowicz, N.; Shah, M. Avaliando Algoritmos de Aprendizagem: Uma Perspectiva de Classificação; Cambridge University Press: Cambridge, Reino Unido, 2011.
52. Boucenna, S.; Gaussier, P.; Andry, P.; Hafemeister, L. Um robô aprende o reconhecimento de expressões faciais e a discriminação face/não face através de um jogo de imitação. Int. J. Soc. Robô. **2014,** 6, 633-652. [CrossRef]
53. Busso, C.; Deng, Z.; Yildirim, S.; Bulut, M.; Lee, CM; Kazemzadeh, A.; Lee, S.; Neumann, U.; Narayanan, S. Análise do reconhecimento de emoções usando expressões faciais, fala e informações multimodais. In Proceedings of the 6th International Conference on Multimodal Interfaces, State College, PA, EUA, 13–15 de outubro de 2004; págs. 205-211.
54. Pantic, M.; Rothkrantz, LJ Rumo a uma interação humano-computador multimodal sensível ao afeto. Proc. IEEE **2003,** 91, 1370-1390. [CrossRef]
55. Cowie, R.; Douglas-Cowie, E.; Tsapatsoulis, N.; Votsis, G.; Kollias, S.; Fellenz, W.; Taylor, JG Emotion reconhecimento na interação humano-computador. Processo de Sinal IEEE. Mag. **2001,** 18, 32-80. [CrossRef]
56. Essa, IA; Pentland, AP Codificação, análise, interpretação e reconhecimento de expressões faciais. Trans. IEEE Padrão Anal. Mach. Intel. **1997,** 19, 757-763. [CrossRef]
57. Mase, K. Reconhecimento da expressão facial do fluxo óptico. IEICE Trans. Inf. Sistema **1991,** 74, 3474-3483.
58. Yacoob, Y.; Davis, L. Computando Representações Espaço-Temporais de Faces Humanas. Ph.D. Tese, Departamento de Ciência da Computação, Universidade de Maryland, College Park, MD, EUA, 1994.

59. Lee, CM; Yildirim, S.; Bulut, M.; Kazemzadeh, A.; Busso, C.; Deng, Z.; Lee, S.; Narayanan, S. Reconhecimento de emoções baseado em classes de fonemas. In Proceedings of the Eighth International Conference on Spoken Language Processing, Jeju Island, Coréia, 4-8 de outubro de 2004.
60. Nwe, TL; Wei, FS; De Silva, LC Classificação das emoções com base na fala. In Proceedings of the IEEE Region 10 International Conference on Electrical and Electronic Technology, TENCON 2001 (Cat. No. 01CH37239), Cingapura, 19–22 de agosto de 2001; Volume 1, pp. 297–301.
61. Mehta, D.; Siddiqui, MFH; Javaid, AY Reconhecimento facial de emoções: uma pesquisa e um usuário do mundo real experiências em realidade mista. Sensores **2018,** 18, 416. [CrossRef]
62. Bharatharaj, J.; Huang, L.; Mohan, R.; Al-Jumaily, A.; Krägeloh, C. Terapia Assistida por Robô para Aprendizagem e Interação Social de Crianças com Transtorno do Espectro do Autismo. Robótica **2017,** 6, 4. [CrossRef]
63. Kosonogov, V.; Zorzi, LD; Honoré, J.; Martínez-Velázquez, ES; Nandrino, JL; Martinez-Selva, JM; Sequeira, H. Variações térmicas faciais: um novo marcador de excitação emocional. PLoS ONE **2017,** 12, e0183592. [CrossRef] [PubMed]
64. Yoshitomi, Y.; Miyaura, T.; Tomita, S.; Kimura, S. Identificação facial usando processamento de imagem térmica. In Proceedings of the 6th IEEE International Workshop on Robot and Human Communication, RO-MAN'97 SENDAI, Sendai, Japão, 29 de setembro a 1 de outubro de 1997; pp. 374-379. [CrossRef]
65. Wang, S.; Ele, M.; Gao, Z.; Ele, S.; Ji, Q. Reconhecimento de emoções de imagens infravermelhas térmicas usando máquina de Boltzmann profunda. Frente. Computar. Sci. **2014,** 8, 609-618. [CrossRef]
66. Bijalwan, V.; Balodhi, M.; Gusain, A. Reconhecimento de emoções humanas usando processamento de imagem térmica e eigenfaces. Int. J. Eng. Sci. Res. **2015,** 5, 34-40.
67. Yoshitomi, Y.; Miyawaki, N.; Tomita, S.; Kimura, S. Reconhecimento de expressão facial usando processamento de imagem térmica e rede neural. In Proceedings of the 6th IEEE International Workshop on Robot and Human Communication, RO-MAN'97 SENDAI, Sendai, Japão, 29 de setembro a 1 de outubro de 1997. [CrossRef]
68. Vukadinovic, D.; Pantic, M. Detecção de Ponto de Característica Facial Totalmente Automática Usando Classificadores Reforçados Baseados em Característica Gabor. In Proceedings of the 2005 IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, HI, EUA, 12 de outubro de 2005.
69. Di Nuovo, A.; Conti, D.; Trubia, G.; Buono, S.; Di Nuovo, S. Sistemas de Aprendizagem Profunda para Estimativa da Atenção Visual na Terapia Assistida por Robô de Crianças com Autismo e Deficiência Intelectual. Robótica **2018,** 7, 25. [CrossRef]
70. Suk, M.; Prabhakaran, B. Sistema de reconhecimento de expressão facial móvel em tempo real—Um estudo de caso. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, EUA, 23–28 de junho de 2014, pp. 132–137.
71. Deshmukh, S.; Patwardhan, M.; Mahajan, A. Pesquisa sobre técnicas de reconhecimento de expressão facial em tempo real. IET Biom. **2016,** 5, 155-163. [CrossRef]
72. Gu, W.; Xiang, C.; Venkatesh, Y.; Huang, D.; Lin, H. Reconhecimento de expressão facial usando codificação radial de características locais de Gabor e síntese de classificador. Reconhecimento de padrões. **2012,** 45, 80-91. [CrossRef]
73. Liu, C.; Wechsler, H. Gabor classificação baseada em recursos usando o modelo discriminante linear de Fisher aprimorado para reconhecimento facial. Trans. IEEE Processo de Imagem. **2002,** 11, 467-476. [PubMed]
74. Boda, R.; Priyadarsini, M.; Pemeena, J. Face detecção e rastreamento usando KLT e Viola Jones. ARPN J. Eng. Aplic. Sci. **2016,** 11, 13472-13476.