

(12) PEDIDO INTERNACIONAL PUBLICADO SOB O TRATADO DE COOPERAÇÃO EM MATÉRIA DE PATENTES
(PCT)

(19) Organização Mundial da
Propriedade Intelectual
Secretaria Internacional

(43) Data de Publicação Internacional
05 de Maio de 2022 (05.05.2022)



(10) Número de Publicação Internacional
WO 2022/087706 A1

(51) Classificação Internacional de Patentes:

G06T 7/11 (2017.01) G06V 10/25 (2022.01)
G06K 9/62 (2022.01)

(21) Número do Pedido Internacional:

PCT/BR2021/050478

(22) Data do Depósito Internacional:

29 de Outubro de 2021 (29.10.2021)

(25) Língua de Depósito Internacional:

Português

(26) Língua de Publicação:

Português

(30) Dados Relativos à Prioridade:

BR 10 2020 022162 0

29 de Outubro de 2020 (29.10.2020) BR

(71) Requerente: **BOTICA COMERCIAL FARMACÊUTI-**

CA LTDA. [BR/BR]; Av. Rui Barbosa, 4110, Parque da
Fonte, 83055-900 São José dos Pinhais (BR).

(72) Inventores: **PADILHA, Milene Haraguchi;** Rua Carlos

Dietzsch, 541, Apto 406 A, Portão, 80330-000 Curitiba
(BR). **URIO, Camila Andréia Bernardon;** Estrada da Gra-
ciosa 3700 casa J12, Alphaville Graciosa, São José dos Pi-
nhais (BR). **SASSON, Clarice Seliar;** Rua Carmelina Ca-
vassin, 1280 casa 4, Curitiba (BR). **DIEAMANT, Gustavo**
De Campos; Rua Haroldo Euclides Souza 140 - Apto 202
torre 1, Curitiba (BR).

(74) Mandatário: **AZEREDO, Fabio José Zanetti de;** Av. Pre-

sidente Juscelino Kubitschek, 1400, 10º andar, 04543-000
Sao Paulo - SP (BR).

(54) Title: METHOD FOR DETECTING AND SEGMENTING THE LIP REGION

(54) Título: MÉTODO DE DETECÇÃO E SEGMENTAÇÃO DA REGIÃO LABIAL

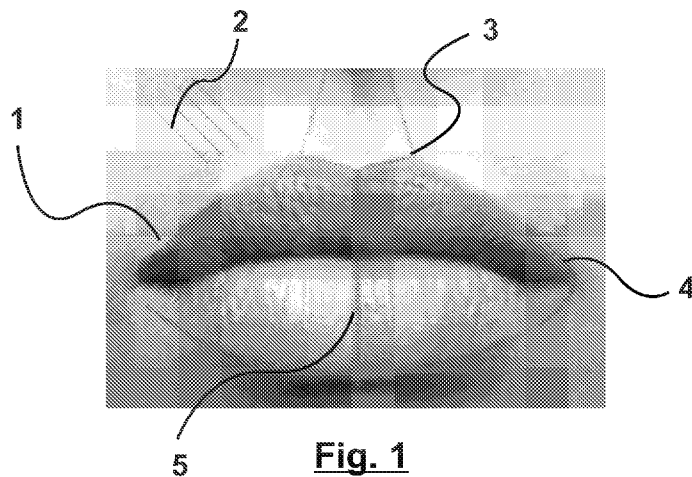


Fig. 1

(57) **Abstract:** The present invention patent pertains to the technical field of methods or arrangements for reading and identifying patterns. More specifically, the present invention relates to technology for using deep learning, machine learning and artificial intelligence algorithms to identify the outline of lips and also to methods for detecting and segmenting the lip region. The method for detecting and segmenting the lip region of the present invention comprises recognizing patterns by extracting input features of lip images, labelling them for training by a processing module in a lip product application system; defining and identifying the labelled lip images with algorithms for recognizing patterns for said lip images to be learnt and segmented by said processing module; and training a machine learning model in said processing module with a plurality of exemplified data and respective answers, defining labels that the model should learn and predict to identify and generate a mathematical pattern for a lip product application system.

(57) **Resumo:** A presente patente de invenção pertence ao campo técnico de métodos ou disposições de leitura e identificação de padrões. Mais especificamente, refere-se à tecnologia de utilização de algoritmos de deep learning, machine learning e inteligência artificial de modo a identificar a delimitação labial bem como métodos que permitam a detecção e a segmentação da região labial. O

(Continua na página seguinte)

(81) Estados Designados (*sem indicação contrária, para todos os tipos de proteção nacional existentes*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Estados Designados (*sem indicação contrária, para todos os tipos de proteção regional existentes*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasiático (AM, AZ, BY, KG, KZ, RU, TJ, TM), Europeu (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarações sob a Regra 4.17:

- *relativa ao direito do requerente de pedir e obter uma patente (Regra 4.17(ii))*
- *relativa à autoria da invenção (Regra 4.17(iv))*

Publicado:

- *com relatório de pesquisa internacional (Art. 21(3))*
- *antes da expiração do prazo para modificar as reivindicações e a republicar na eventualidade de receção de tais modificações (Regra 48.2(h))*
- *em preto e branco; o pedido internacional tal como depositado contém cores ou níveis de cinza e pode ser baixado do PATENTSCOPE*

método de detecção e segmentação da região labial da presente invenção compreende a realização do reconhecimento de padrões por meio da extração de características de entrada de imagens labiais, rotulando-as para uma base de treinamento por meio de um módulo de processamento presente em um sistema de aplicação de produtos labiais; definição e indicação das imagens labiais rotuladas aos algoritmos de reconhecimento de padrões para aprendizagem e realização de segmentação das ditas imagens labiais pelo dito módulo de processamento; e a execução de treinamento de um modelo em aprendizagem de máquina no dito módulo de processamento com uma pluralidade de dados exemplificados e suas respectivas respostas definindo rótulos que o modelo deve aprender e prever para identificar e gerar um padrão matemático para um sistema de aplicação de produtos labiais.

"MÉTODO DE DETECÇÃO E SEGMENTAÇÃO DA REGIÃO LABIAL"

[001] A presente patente de invenção pertence ao campo técnico de métodos ou disposições de leitura e identificação de padrões. Mais especificamente, refere-se à tecnologia de utilização de algoritmos de *deep learning*, *machine learning* e inteligência artificial de modo a identificar a delimitação labial bem como métodos que permitam a detecção e a segmentação da região labial.

ESTADO DA TÉCNICA

[002] O estudo do estado da técnica é extremamente importante para o resultado proposto pela presente invenção, visto que não foi encontrada nenhuma solução que cumprisse os requisitos de dispor de um processo de detecção e reconhecimento de imagens labiais precisos, por meio do uso de inteligência artificial, visando a utilização por um sistema inteligente de aplicação de produtos labiais nos lábios com precisão.

[003] Neste sentido, é realizado um levantamento das tecnologias relacionadas, direta ou indiretamente, com os problemas vinculados ao objetivo da presente invenção, tais como: segmentação dos lábios, reconhecimento de bordas e texturas, aplicação de pigmentos, entre outros. Também é realizada uma análise dos dispositivos existentes no mercado e que estão relacionados com alguma parte do processo de aplicação de produtos de beleza.

[004] Diversos documentos do estado da técnica mencionam a utilização de *deep learning*, *machine learning* e inteligência artificial visando a identificação facial, por meio de processos de detecção e reconhecimento de imagens previamente captadas, ou, ainda, em imagens em tempo real.

[005] O documento de patente BR102012033722-3 apresenta

um sistema e método capaz de realizar maquiagem virtual em imagens obtidas por dispositivos portáteis equipados com câmeras digitais. De acordo com o pedido de patente, o método permite a identificação automática de pontos de interesse (olhos, boca, sobrancelha, contorno do rosto) na imagem. Outro documento de patente, EP2985732B1, apresenta um processamento de imagem novamente com a finalidade de reprodução de maquiagem virtual. Já o documento de patente US10339685B2 apresenta um sistema e um método capazes de remover digitalmente uma maquiagem facial. Todos esses documentos apresentam aplicação de tecnologias de reconhecimento facial que consideram a face como um todo. Essas metodologias correlacionam diversos pontos da face, como olhos, sobrancelhas, nariz e boca, de forma a facilitar o reconhecimento de cada uma dessas regiões.

[006] O documento de patente US9552510B2 revela técnicas para captura de expressões faciais para desenhos animados por meio de inteligência artificial, porém, novamente, correlaciona pontos chave da face como um todo para obter determinado reconhecimento facial.

[007] Outra aplicação da tecnologia de inteligência artificial no campo técnico de cosméticos está relacionada a fornecer recomendações para produtos customizados, conforme demonstrado no documento de patente WO2017/165363A1. Dito documento revela o reconhecimento da idade da pele por meio de imagens capturadas por dispositivos. Após analisadas por dispositivos computacionais, a idade da pele é determinada e formulações específicas são sugeridas para cada usuário.

[008] Quanto a métodos de reconhecimento, avaliação e categorização de lábios, sem a aplicação da tecnologia de

inteligência artificial, podem ser citados documentos de patente como WO2006043643A1 e WO2008135907A1.

[009] Apesar de não pertencer ao campo de aplicação de cosméticos, o documento não-patentário intitulado "LipNet: End-to-end sentence-level Lipreading" revela a utilização de inteligência artificial, por meio de redes, para a realização de leitura labial com alta precisão. Ainda que seja uma aplicação focada em lábios, está diretamente relacionada aos seus movimentos sequenciais, bem como ao contexto de vocabulários e sentenças, que fazem parte do aprendizado da leitura.

[0010] Outros diversos documentos de literatura não patentária são conhecidos no estado da técnica, tendo como objetivo a detecção e segmentação dos lábios em imagens coloridas.

[0011] A metodologia descrita por Eveno, N. et al (Eveno, N., Caplier, A., & Coulon, P. Y. (2002, August). Key points based segmentation of lips. In Proceedings. IEEE International Conference on Multimedia and Expo (Vol. 2, pp. 125-128). IEEE) é baseada na detecção de Key Points labiais, que sinalizam pontos ou características representativas de uma região em uma imagem, a fim de estimar o contorno externo dos lábios. Nesta abordagem são utilizadas técnicas de processamento digital de imagens no espaço de cor RGB, e a análise da intensidade luminosa incidente sobre o lábio, para a identificação de lábio superior e inferior, uma vez que a iluminação incide no lábio superior de maneira diferente do lábio inferior. Neste documento, os contornos labiais são definidos através do rastreamento da coloração ao redor de cada pixel na imagem. Os contornos encontrados são

considerados como curvas polinomiais, em que picos e vales definem as regiões do arco do cupido (estrutura da anatomia labial localizada no centro do lábio superior, sendo uma depressão em forma de V localizada logo acima do *vermilion*) e as extremidades das comissuras (nome que se dá as junções dos bordos de aberturas em forma de fenda, no corpo (pálpebras, lábios etc.)). Os testes são realizados com imagens da região labial, com pessoas falando sob iluminação natural e não uniforme. O algoritmo desenvolvido mostrou-se robusto nos casos em que o indivíduo possuía barba, e quando os dentes e língua estavam expostos. Por outro lado, os problemas para essa abordagem estão relacionados à rotação do rosto do indivíduo, fazendo com que a visibilidade e a definição das características labiais sejam perdidas quando há a mudança de ângulo do rosto, e a definição apenas das bordas externas e não internas dos lábios.

[0012] O documento Hassanat, A. B. et al (Hassanat, A. B. & Jassim, S. (2010, April). Color-based lip localization method. In Mobile Multimedia/Image Processing, Security, and Applications 2010 (Vol. 7708, p. 77080Y). International Society for Optics and Photonics) faz a análise da frequência da intensidade de cor presente nos lábios e da face a fim de identificar as regiões de interesse. A abordagem da diferenciação entre pele dos lábios e da face são feitas a partir da análise da frequência da intensidade de cor nos espaços de cor: RGB, HSV e YCbCr. Para uma aproximação mais precisa da região labial é aplicado um algoritmo de clusterização (agrupamento por características semelhantes pré definidas - técnica de agrupamento de dados correlacionados dado um espaço n-dimensional) por pixels de

intensidade de cor e vizinhança. Este documento apresenta uma aplicação em vídeos para voluntários de diferentes etnias (africanos, asiáticos, europeus, pessoas do oriente médio), do sexo feminino e masculino, e voluntários com barba e/ou bigode. O modelo proposto por Hassanat, A. B. et al tem como desafio a utilização de espaços de cores para a segmentação labial, uma vez que a intensidade luminosa é crucial para que as características labiais de cor sejam reconhecidas pelo método proposto. O método obtém uma acurácia média de 91,15 % em relação à localização da região labial. Por outro lado, os piores resultados para essa abordagem estão relacionados à análise de imagens no espaço de cor RGB. É verificada uma robustez no algoritmo quando há a detecção dos lábios em indivíduos do sexo feminino, uma vez que indivíduos do sexo masculino possuem pelos na região facial, fato que acaba confundindo o algoritmo.

[0013] Em Skodras, E. et al (Skodras, E., & Fakotakis, N. (2011, May). An unconstrained method for lip detection in color images. In 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1013-1016). IEEE), a metodologia utilizada baseia-se na segmentação por cor, a partir da imagem da face, para obter a melhor região que representa a separação entre lábios e a pele do rosto. A região é baseada na diferença de cor entre lábios e pele e essa separação é independente das intensidades particulares dos pixels da partição. Para esse propósito, o espaço de cor utilizado é o $L^*a^*b^*$ e as técnicas combinadas de segmentação de cor utilizando Nearest Neighbor (análise de vizinhança através da proximidade dos dados em um espaço n-dimensional) e Clusterização baseada em cores

utilizando K-Means (algoritmo de clusterização que agrupa dados correlacionados de acordo com determinado raio de vizinhança) são aplicadas. Posteriormente, um processamento matemático é aplicado, uma elipse é ajustada da melhor forma aos lábios detectados e assim os pontos de interesse (Key Points) referentes aos lábios são extraídos. Os testes apresentados por Skodras, E. et al são realizados com imagens de bases utilizadas sob diversas condições de iluminação e não há nenhum tipo de maquiagem sendo utilizada pelas pessoas presentes nas imagens. Além disso, imagens com baixa iluminação também são utilizadas, o que dificulta a segmentação por cores. O método é robusto mesmo em casos desafiadores, como em situações em que a luz não é uniforme, presença de barba, baixo contraste de cores entre a área labial e não labial ou se os dentes são visíveis. Também não é afetado pela inclinação e rotação da cabeça, desde que a região dos lábios seja visível. Por outro lado, a abordagem de Skodras, E. et al falha em extrair resultados precisos nos casos em que os cantos da boca foram cobertos por partes da barba.

[0014] O documento Le, T. H. N et al (Le, T. H. N., & Savvides, M. (2016). A novel shape constrained feature-based active contour model for lips/mouth segmentation in the wild. *Pattern Recognition*, 54, 23-33) utiliza a abordagem automática de detecção e segmentação dos lábios/boca sem restrição de condições do ambiente e robusto à posição do rosto e oclusão. Primeiramente realiza-se a predição de Key Points que representam os pontos mais importantes dos lábios. Esses Key Points são utilizados posteriormente para realizar a segmentação do objeto de interesse: os lábios/boca. Para a

predição dos Key Points utiliza-se um modelo baseado na técnica Active Shape Model. Após isso, é proposto o modelo SC-FAC (Shape Constrained Feature-based Active Contour) que realiza a segmentação dos lábios/boca com base nos Key Points preditos. Os testes em Le, T. H. N et al são realizados com imagens de diversas bases, sendo que as imagens utilizadas estão sob condições de iluminação distintas, baixo contraste, pessoas com expressões faciais, baixa resolução, embaçadas e maquiagem. O método proposto é robusto às condições testadas apresentando bons resultados.

[0015] Diferentemente das abordagens anteriores utilizadas que se baseiam na extração dos Key Points para posteriormente realizar a segmentação, o trabalho apresentado em Dong, W et al (Dong, W., He, R., & Zhang, S. (2016, October). Digital recognition from lip texture analysis. In 2016 IEEE International Conference on Digital Signal Processing (DSP) (pp. 477-481). IEEE) propõe uma abordagem de reconhecimento de lábios através da textura utilizando Deep Learning. A arquitetura proposta é composta por redes do tipo Convolutional Neural Network (CNN) (tipo de rede neural artificial utilizado no reconhecimento de padrões) e Long Short-Term Memory (LSTM) para reconhecimento digital, que integra as vantagens da CNN e LSTM. Utilizou-se a CNN com o objetivo de aprender recursos avançados de imagem para aparência dos lábios e LSTM para aprender a variação espaço-temporal entre frames (cada uma das imagens que, quando postas em sequência, dão origem ao vídeo). Por outro lado, o método proposto não apresenta robustez à variação de iluminação.

[0016] Por sua vez, o objetivo do trabalho proposto por

Bindal, U et al (Bindal, U., Shajan, K., Mitra, N. K., & Priyadarshni, B. (2015). Morphological Analysis of Vermilion Border and its Forensic Applications. J Dentistry Oral Medicine, 3 (1), 21-28) é encontrar as características mais comuns presentes nos lábios no que diz respeito às variações sexuais, à morfologia dos lábios e suas medidas antropométricas. Os experimentos realizados em Bindal, U. et al contaram com a participação de 300 indivíduos na faixa etária entre 18 e 65 anos. As impressões da borda dos lábios foram obtidas e a análise morfométrica das impressões obtidas foram feitas usando uma régua. A altura e largura da boca foram medidas e as rugas e sulcos nas impressões obtidas foram estudadas com a ajuda de uma lente de aumento. As impressões de um indivíduo foram repetidas a cada ano durante sete anos. Através dos experimentos constatou-se que a largura da boca e a espessura dos lábios nos homens, geralmente, são maiores do que nas mulheres. Foi constatado também que morfologicamente lábios médios estão presentes em ambos os sexos e mudanças nas medidas antropométricas não foram observadas nos sete anos de estudos posteriores.

[0017] O documento apresentado em Liu, Y. et al (Liu, Y., Li, Y., Ma, X., & Song, R. (2017). Facial expression recognition with fusion features extracted from salient facial areas. Sensors, 17 (4), 712) tem como foco principal o reconhecimento de expressões faciais. A abordagem proposta consiste em um algoritmo que utiliza a fusão de características extraídas das áreas salientes das faces. Para extrair características mais representativas, primeiro define-se as áreas salientes nas faces da imagem e essas áreas são utilizadas para extrair características em outras

imagens de faces de tamanhos similares. Portanto, pode-se extrair características mais semelhantes de diferentes indivíduos. As características provenientes de técnicas de segmentação por textura de uma imagem (Local Binary Pattern - LBP) e de descritores de características de detecção de objetos (Histogram of Oriented Gradient - HOG) são extraídas das áreas salientes, a dimensão das fusões de características são reduzidas pelo Principal Component Analysis (PCA) (procedimento matemático que converte um conjunto de variáveis correlacionadas em linearmente não correlacionadas, chamadas de Componentes Principais) e aplica-se vários classificadores (modelos de Reconhecimento de Padrões) para classificar as seis expressões básicas (alegria, medo, surpresa, tristeza, nojo, raiva) de uma só vez. Como resultado, as áreas salientes encontradas em diferentes indivíduos são do mesmo tamanho. Além disso, a correção da iluminação da imagem é aplicada primeiramente aos recursos de LBP na estrutura de algoritmos visando melhorar as taxas de reconhecimento. Diversas bases de imagens são utilizadas nos experimentos realizados em Liu, Y. et al. Apesar do método proposto em Liu, Y. et al ter alcançado o estado da arte no que diz respeito ao reconhecimento de expressões faciais, há dois pontos em que a abordagem não apresenta boa performance. O primeiro ponto, trata-se do tamanho da base de imagens utilizada. Se a base de imagens não for suficientemente grande, o método de correção de luminosidade utilizado não consegue melhorar o reconhecimento das expressões. O segundo ponto, trata-se da utilização de Landmarks (importantes pontos de interesse dos objetos na imagem utilizados como referência pelo modelo de reconhecimento de padrões), os

quais são pontos que representam as principais características presentes na face. Se esses pontos não forem precisos, o reconhecimento das expressões faciais também não será.

[0018] O trabalho proposto por He, K. et al (He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969)) propõe uma arquitetura chamada Mask R-CNN que realiza o processo de segmentação, abordado em outros artigos já citados, porém, baseada em Deep Learning. O Mask R-CNN é um método de segmentação de objetos, baseado no Faster R-CNN (consiste em uma arquitetura de Deep Learning para realização da detecção de objetos), que inicialmente necessita realizar a detecção dos objetos presentes na imagem para em seguida realizar a segmentação de tais objetos. Então, em um primeiro momento a arquitetura da rede do Faster R-CNN é utilizada para realizar a detecção dos objetos e após isso com a informação da detecção, de forma paralela, é feito a segmentação dos objetos através da criação de máscaras. A abordagem proposta em He, K. et al supera os trabalhos anteriores do estado da arte e consegue fornecer bons resultados ainda sob condições desafiadoras.

[0019] Diferentemente dos trabalhos já citados, o trabalho de Kobayashi, H. et al (Kobayashi, H., & Tagami, H. (2004). Functional properties of the surface of the vermilion border of the lips are distinct from those of the facial skin. British Journal of Dermatology , 150 (3), 563-567) propõe apresentar as diferenças entre a pele do lábio e da face através das propriedades funcionais. O estudo de Kobayashi, H. utilizou um sistema de câmaras para medição da

perda transepitelial de água (TEWL) do lábio. A partir disso foi feito um estudo das propriedades biofísicas do lábio comparando-as com as da pele da bochecha. Os testes de TEWL, condutância de alta frequência, são realizados com 303 voluntários saudáveis para a região labial e região da bochecha. Os dados obtidos mostram que o lábio possui três vezes mais capacidade de perda de água em relação a região da bochecha. Apesar de apresentar diferenciação de dados obtidos pela TEWL, o ambiente experimental é controlado em condições de umidade (50 %) e temperatura (21 °C).

[0020] Conforme demonstrado, é possível verificar que a tecnologia de inteligência artificial vem sendo aplicada em diversas frentes relacionadas ao reconhecimento facial.

[0021] Quando aplicada especificamente aos lábios, entende-se que a maioria das técnicas faz referências a outros pontos do rosto para que a identificação dos lábios seja possível. Quando focados unicamente na boca, ainda assim necessitam pontos de referência (pontos chave), até mesmo como dentes ou barbas, para que a identificação dos lábios se torne mais assertiva. Além dos pontos chave e, portanto, a necessidade de visualização da boca completa, ainda existem outros fatores limitantes, como iluminação e ângulo dos lábios.

[0022] Além disso, a borda labial não é uma fronteira bem definida e, além de possuir uma variação não linear de tonalidade, pode possuir assimetrias no traçado. As técnicas para reconhecimento labial identificadas no estado da técnica apresentam problemas relacionados à precisão para identificação da borda que define o limite dos lábios.

[0023] Conforme verificado, foram localizados diversos

documentos que se remetem ao campo técnico da presente invenção, contudo, não se direcionam a aperfeiçoar etapas de processamento de imagens necessárias para distinguir com extrema precisão os limites entre as áreas da derme facial e labial de modo a possibilitar a aplicação de tecnologia em hardware com relação a um sistema inteligente de aplicação de produtos labiais nos lábios.

[0024] A presente invenção possui como principal objetivo a utilização de métodos de Inteligência Artificial que permitem a identificação e segmentação dos lábios de forma precisa, para que seja possível a aplicação automática de produtos labiais sem a utilização de acessórios, como espelhos.

BREVE DESCRIÇÃO DAS FIGURAS

[0025] A descrição que se segue não se limita aos desenhos ou componentes citados, tendo como referência as seguintes ilustrações referenciadas abaixo.

[0026] A Figura 1 apresenta uma imagem representativa de identificação das diversas regiões de um contorno labial.

[0027] A Figura 2 apresenta um gráfico das frequências captadas pelos cones no olho humano.

[0028] A Figura 3 apresenta um gráfico com a representação do espaço de cor RGB.

[0029] A Figura 4 apresenta um gráfico de comparação de três espaços de cores (Lab, RGB e CMYK).

[0030] A Figura 5 apresenta uma imagem de representação do espaço de cor Lab.

[0031] A Figura 6 apresenta uma imagem de representação do espaço de cor HSV.

[0032] A Figura 7 apresenta uma imagem de comparação

espacial dos espaços de cores YCbCr e o espaço RGB.

[0033] A Figura 8 apresenta uma detecção de objetos/elementos na imagem e seus respectivos bounding boxes em vermelho, verde e azul.

[0034] A Figura 9 apresenta um reconhecimento de objetos/elementos na imagem e seus respectivos rótulos.

[0035] A Figura 10 apresenta uma segmentação de objetos/elementos na imagem e suas respectivas máscaras em vermelho, verde e azul.

[0036] A Figura 11 apresenta um diagrama do processo de treinamento em aprendizagem de máquina.

[0037] A Figura 12 apresenta um diagrama do processo de treinamento em aprendizagem profunda.

[0038] As Figuras 13A, 13B e 13C apresentam exemplos de imagens adquiridas usando câmera infravermelha do Leap Motion.

[0039] A Figura 14 apresenta imagens de alta resolução obtidas por uma câmera webcam modificada para ter a menor configuração de foco possível e capaz de capturar imagens em full HD.

[0040] As Figuras 15A, 15B, 15C e 15D apresentam a utilização de imagens de microscopia.

[0041] A Figura 16 apresenta imagens de alta definição das curvas e desenho dos lábios na aplicação do pigmento de batom proposta pela presente invenção.

[0042] A Figura 17 apresenta exemplos de imagens do dataset SCface - Surveillance Cameras Face Database composta por imagens da face capturadas por câmeras distintas, uma RGB e outra infravermelha.

[0043] As Figuras 18A e 18B apresentam resultados de

predição de contornos da U-Net.

[0044] A Figura 19 apresenta um diagrama do processo de treino e teste da rede U-Net utilizado para segmentação da boca.

[0045] A Figura 20 apresenta resultados da rede U-Net para a predição dos contornos da boca.

[0046] A Figura 21 apresenta uma imagem utilizada por uma ferramenta de rotulação de imagens de treinamento.

[0047] A Figura 22 apresenta um diagrama do processo de treino e teste aplicados para a rede U-Net com variações da máscara de predição.

[0048] A Figura 23 apresenta exemplos de resultados para as diferentes máscaras utilizando a rede U-Net.

[0049] A Figura 24 apresenta exemplos de resultados da rede U-Net com a máscara de contorno labial mais espessa.

[0050] As Figuras 25A e 25B representam a detecção de objetos e a segmentação dos mesmos objetos com as respectivas máscaras.

[0051] As Figuras 26A à 26F apresentam exemplos de como é feita a base de imagens para treinamento utilizando apenas partes dos lábios.

[0052] A Figura 27 apresenta um diagrama do fluxo de execução do treinamento e teste na realização da segmentação da região dos lábios em um primeiro cenário de acordo com a presente invenção.

[0053] A Figura 28 apresenta um diagrama do fluxo de execução do treinamento e teste na realização da segmentação da região dos lábios em um segundo cenário de acordo com a presente invenção.

[0054] A Figura 29 apresenta um diagrama do fluxo de

execução do treinamento e teste na realização da segmentação da região dos lábios em um terceiro cenário de acordo com a presente invenção.

[0055] A Figura 30 apresenta imagens de resultados satisfatórios do processo de segmentação dos lábios de acordo com o primeiro cenário.

[0056] A Figura 31 apresenta imagens de resultados satisfatórios do processo de segmentação dos lábios de acordo com o segundo cenário.

[0057] A Figura 32 apresenta imagens de resultados satisfatórios do processo de segmentação dos lábios de acordo com o terceiro cenário.

[0058] A Figura 33 apresenta imagens de resultados insatisfatórios do processo de segmentação dos lábios de acordo com o terceiro cenário.

[0059] A Figura 34 ilustra um espaço de cor de uma imagem digital (RGB) e o agrupamento realizado.

[0060] As Figuras 35A e 35B apresentam imagens resultantes da aplicação do método de superpixel sobre uma imagem digital e a resultante da clusterização por cor.

[0061] As Figuras 36A à 36E apresentam imagens de zoneamento por superpixel aplicado em borda de lábio sem pigmento.

[0062] As Figuras 37A à 37E apresentam imagens de zoneamento por superpixel aplicado em borda de lábio com pigmento.

[0063] As Figuras 38A à 38E apresentam imagens de zoneamento por superpixel aplicado em borda de lábio sem pigmento.

[0064] As Figuras 39A à 39C apresentam imagens de

zoneamento por superpixel aplicado em borda de lábio sem pigmento no espaço HSV.

[0065] As Figuras 40A à 40C apresentam imagens de zoneamento por superpixel aplicado em borda de lábio sem pigmento no espaço HSV e em RGB.

[0066] A Figura 41 apresenta um diagrama de fluxo de execução da junção das bordas de superpixel e Mask R-CNN.

[0067] A Figura 42 apresenta imagens de resultados satisfatórios do processo de segmentação dos lábios de acordo com a execução da junção das bordas de superpixel e Mask R-CNN em um primeiro cenário.

[0068] A Figura 43 apresenta imagens de resultados satisfatórios do processo de segmentação dos lábios de acordo com a execução da junção das bordas de superpixel e Mask R-CNN em um segundo cenário.

DESCRIÇÃO DETALHADA

[0069] A presente invenção se refere a um método de estabelecimento de informações labiais por meio de um módulo de processamento capaz de determinar e suportar um sistema de aplicação de produtos labiais nos lábios com precisão, mais especificamente apresenta um método de delimitação labial com etapas que permitam a detecção e a segmentação da região labial.

[0070] Inicialmente faz-se necessário identificar os lábios e separar estes do restante da pele do rosto. Como se trata de um processo que visa a aplicação posterior de produtos labiais nos lábios com precisão, é necessário ter informações da forma e do contorno da boca, sendo isto um dos fatores determinantes para a aplicação precisa do batom.

[0071] A principal motivação que propicia o processo de

estabelecimento de informações labiais da presente invenção se refere a aplicação de tecnologia em hardware com relação a um sistema inteligente de aplicação de produtos labiais nos lábios com precisão por meio de etapas de processamento necessárias para distinguir com precisão o limite entre as áreas da derme facial e labial.

[0072] Partindo da análise de conceitos e definições para a viabilidade de um processo de estabelecimento de informações labiais que consiga cumprir com o desafio de aplicar produtos labiais por meio de um sistema inteligente de aplicação de produtos labiais nos lábios com precisão, algumas definições e conceitos são aqui previamente estabelecidos.

[0073] Existem três áreas anatômicas reconhecidas nos lábios, como relata a introdução do artigo Bindal, U. et al (Bindal, U., Shajan, K., Mitra, N. K., & Priydarshni, B. (2015). Morphological Analysis of Vermilion Border and its Forensic Applications. J Dentistry Oral Medicine, 3 (1), 21-28): pele externa, vertente livre (área vermelha, externa a boca, que é considerada uma semi mucosa) e vertente interna (mucosa oral). No que se refere à borda labial, se apresenta como sendo a fronteira da epiderme que separa a pele externa, altamente queratinizada, da vertente livre que é menos queratinizada, enquanto os lábios, se apresentam como sendo a vertente livre.

[0074] A pele do rosto é mais espessa (dezesseis camadas celulares) do que a pele dos lábios (três a cinco camadas celulares) que são compostos inteiramente por tecidos moles e assim possuem os vasos sanguíneos mais próximos à superfície. Ademais, a pele dos lábios por não possuir folículos pilosos

e glândulas sudoríparas e sebáceas, não possui a camada de proteção usual de suor e óleos corporais que mantêm a pele suave, inibem patógenos e regulam o calor. Por esse motivo, os lábios secam mais rápido e ficam mais rachados.

[0075] Ainda em relação à região labial, a Figura 1 ilustra a denominação das diversas partes de uma imagem relacionadas à definição do contorno labial (1), tais como linhas periorais (2), filtro e arco do cupido (3), comissura labial (4) e proporção/equilíbrio (5).

[0076] As técnicas para reconhecimento labial usuais possuem problemas relacionados à precisão para identificação da borda que define o limite dos lábios, consequentemente carecem de melhor definição da área de aplicação do pigmento de um produto labial. A borda labial não é uma fronteira bem definida e além de possuir uma variação não linear de tonalidade, pode possuir assimetrias no traçado. Então é necessário entender como os usuários de produtos labiais definem essa borda.

[0077] Normalmente os usuários costumam utilizar produtos de pigmentação nos lábios seguindo o desenho da sua borda labial. Muitas vezes essa aplicação costuma ultrapassar de forma sutil a borda labial em certos pontos. Mas isso não compromete a qualidade da aplicação do pigmento, ao contrário, parece ser necessário para criar um aspecto mais simétrico e perfeito dos lábios.

[0078] Uma classe de produtos labiais, os batons, podem ser usados para obtenção do efeito de aumento dos lábios. Baseando-se na ideia de aplicar o produto além da borda dos lábios, pigmentando parte da pele próxima. A definição de onde será essa nova borda labial, parte de questões estéticas

para o desenho de uma boca bem delineada.

[0079] No que se refere a identificação de espaços de cor, as cores visíveis pelo sistema visual humano, são conhecidas por estarem dentro de um espectro de luz, onde ao variar a frequência da luz observamos uma variação da cor. O olho humano por sua vez, é composto por diversas estruturas que o permitem focalizar, alterar a intensidade de luz e identificar essas frequências de luz visível. Sendo que o processo de captura desta luz é feito por células denominadas cones, já a captura de luminosidade (claro ou escuro) é feito por células denominadas bastonetes. O olho humano é composto, em sua maioria, por três tipos de cones, os cones de baixa frequência que são mais sensíveis a frequências de luz azul, de média frequência que são mais sensíveis a frequências de luz verde e os cones de alta frequência que são mais sensíveis a frequências de luz vermelha. O fato de um tipo de cone ser mais sensível a uma frequência do que outras, não significa que ele não consiga captar essas outras frequências, desta forma existe uma intersecção nas frequências captadas, o que permite ao olho humano, em geral, observar todo o espectro de luz visível, essa intersecção pode ser observada na Figura 2 .

[0080] Considerando essas frequências visíveis, diversos modelos matemáticos foram desenvolvidos ao longo do tempo para representar essa variação de cores, estes modelos são chamados em processamento de imagens de espaços de cores. O espaço de cor mais conhecido e mais utilizado é o RGB. Este espaço é representado por um cubo onde as cores variam nos eixos, sendo o ponto de origem (0,0,0) a representação da cor preta e seu extremo oposto o ponto final (255,255,255) a

representação da cor branca, sendo cada eixo representando as cores R - vermelho, G - verde e B - azul, como ilustra a Figura 3.

[0081] Esse espaço é extremamente utilizado em equipamentos de captura e de reprodução de imagens como: câmeras fotográficas, monitores, televisores, projetores, entre outros. Apesar deste modelo representar uma grande quantidade de cores e intensidades, o mesmo não consegue representar todas as cores visíveis pelo olho humano, nem permite um controle mais sutil de informações como luminosidade, saturação e transparência, como é possível observar na comparação entre o espaço de cor Lab (6), o espaço de cor RGB (7) e o espaço de cor CMYK (8) apresentados na Figura 4.

[0082] O espaço de cor Lab ou CieLab é considerado o espaço de cor que mais se aproxima da quantidade de cores visíveis, tanto em quantidade de cores, quanto na intensidade luminosa. Esse espaço de cor consiste em três eixos, assim como no RGB, porém ao invés de ter um ponto inicial (0,0,0) este espaço tem coordenadas que variam dentro de um espectro, sendo o L a luminância normalmente variando de 0 até 100 ou do preto até o branco, o A e o B variando de -128 a +128 o que respectivamente significa que variam do verde até o vermelho e do azul até o amarelo, como mostra a Figura 5. A grande vantagem desse espaço em relação ao RGB, além da possibilidade de se ter uma quantidade maior de cores representadas é a possibilidade de se trabalhar com o canal da luminância (L) de forma separada, podendo assim modificar algumas características da imagem sem modificar a informação de cor da imagem.

[0083] Outro espaço de cor muito utilizado em processamento de imagens é o HSV - Hue, Saturation, Value. Este espaço de cor é representado no espaço como um cilindro, conforme visto na Figura 6, onde a saturação (S) é o raio, o valor (V) é a altura e o espectro de cor (H) é o perímetro, desta forma o branco é representado pela parte central do topo do cilindro, enquanto o preto é representado pela base do cilindro. Esse espaço de cor é muito utilizado em aplicações que envolvem operações com a pele humana, já que ele permite de forma fácil, identificar diferenças sutis de tonalidades, além de separar a informação cor, das outras e assim como o espaço de cor Lab, permite realizar operações nos canais, sem perder a informação de cor.

[0084] Por fim, outro espaço de cor utilizado ao longo desse projeto é o espaço de cor YCbCr onde o Y é a luminância, enquanto o Cb e o Cr são respectivamente a diferença de azul e a diferença de vermelho. A representação desse espaço de cor é um cubo, assim como o espaço de cor RGB, e pode ser considerado uma extensão do espaço de cor RGB, contudo esse espaço de cor contém uma faixa de cores muito maior que a faixa de cores do espaço de cor RGB, e por possuir uma separação de canais, onde o Y é a luminância, permite uma quantidade distinta de operações. A Figura 7 mostra a comparação espacial entre os espaços de cores YCbCr (cubo externo) e o RGB (cubo interno).

[0085] A importância de se processar as imagens em espaços distintos está diretamente relacionada com as diferentes operações matemáticas que cada espaço permite. Desta forma uma informação que em um espaço de cor é misturado com outras informações, em outro espaço essa

informação está independente, permitindo ajustes e calibrações específicas.

[0086] No que se refere à detecção, segmentação e reconhecimento de objetos em imagens digitais, a visão computacional é o campo da computação que estuda maneiras de extrair informações dos objetos de uma imagem. Atua juntamente com o processamento digital de imagens, analisando imagens para obter um resultado próximo do olhar humano. Em outras palavras, a visão computacional é o processo de modelagem e replicação da visão humana utilizando recursos de hardware implementados por programa computacional, compreendendo uma cena 3D a partir de imagens 2D em termos das propriedades das estruturas presentes na cena. O reconhecimento de objetos é uma das principais funções ligadas à visão computacional e está intimamente ligado ao reconhecimento de padrões. Um objeto pode ser definido por seus padrões de textura, cor, forma, dimensão, entre outros, e o reconhecimento individual destes padrões podem caracterizar um objeto como um todo, ou seja, ser definido como uma classe. Uma classe de padrões é uma família de padrões que compartilham entre si propriedades em comum.

[0087] O reconhecimento de padrões por equipamentos envolve técnicas de atribuição de padrões às suas respectivas classes de forma automática e com a menor intervenção humana possível, um exemplo disto é a utilização de algoritmos de aprendizagem de máquina para a classificação de objetos. Ao extrair informações de uma imagem, é necessário analisar tais propriedades com o auxílio dos algoritmos de aprendizado de máquina, estimando um padrão dado um novo exemplo, a partir do aprendizado previamente adquirido. Entretanto o

aprendizado de máquina sofre limitações, pois ao determinar a qual classe um objeto pertence é necessário a extração da melhor representação possível, e isso envolve um trabalho cauteloso nas áreas de engenharia.

[0088] Em visão computacional existem técnicas para a detecção e a segmentação de objetos na imagem. A detecção de objetos, por sua vez, consiste em localizar os objetos de interesse presentes na imagem e após essa localização, para cada objeto é construído um bounding box (que consiste em uma caixa retangular que pode ser determinada pelas coordenadas dos eixos x e y no canto superior esquerdo e inferior direito do retângulo) correspondendo à localização daquele objeto na imagem, tal como representado na Figura 8. Porém, o bounding box apenas informa que há um elemento naquela região, mas não fornece nenhum tipo de informação adicional como forma, contorno e a identificação do objeto.

[0089] O reconhecimento é uma etapa que consiste na identificação dos objetos sempre posterior à etapa de detecção. Um dos requisitos desta etapa é a informação do objeto, ou seja, a região da imagem onde encontra-se o objeto que o sistema deverá analisar para realizar o processo de reconhecimento utilizando um modelo de aprendizagem de máquina previamente treinado. A Figura 9 exemplifica a detecção e a identificação dos objetos da imagem de acordo com seus respectivos rótulos (gato e cachorro).

[0090] Por fim, a segmentação consiste em criar uma máscara para cada objeto presente na imagem, de forma a particionar a imagem em segmentos relevantes. Essas máscaras fornecem informações mais granulares sobre o objeto como localização, topografia e a forma. A Figura 10, representa a

segmentação de três objetos distintos por meio de suas respectivas máscaras (em vermelho, verde e azul).

[0091] No que se refere a Aprendizagem de Máquina (Machine Learning), é uma subárea da Inteligência Artificial (IA) que consiste na criação de modelos que são capazes de identificar padrões matemáticos gerando saídas baseadas nas informações de entrada. Para o modelo ser capaz de identificar um padrão matemático, é necessário que esse sistema seja treinado previamente com inúmeros exemplos de dados e suas respectivas respostas (rótulos, que representam valores verdadeiros e falsos que um modelo deve aprender e prever)

[0092] No processo de treinamento, conforme apresentado no diagrama da Figura 11, a partir das informações de entrada, algumas características serão extraídas e em seguida, juntamente com seus rótulos, são submetidas ao processo de treinamento do modelo. Ao final do processo de treinamento, o modelo gerado fornecerá uma saída capaz de identificar o rótulo de um objeto.

[0093] Por sua vez, a Aprendizagem Profunda (Deep Learning) é uma sub área da Aprendizagem de Máquina. A principal diferença entre as técnicas de Aprendizagem de Máquina e as baseadas em Aprendizagem Profunda é a complexidade dos algoritmos, tanto nos modelos matemáticos utilizados quanto no número de etapas adicionais presentes neste modelo. Essas etapas adicionais e mais complexas permitem que o modelo consiga generalizar as informações, conseguindo extrair características e detalhes dos dados de entrada. Uma desvantagem da abordagem que utiliza a Aprendizagem Profunda é a necessidade de uma quantidade ainda

maior de dados de entrada, bem como uma variação desses dados. Um algoritmo de Aprendizagem de Máquina consegue convergir utilizando centenas de dados de entrada, enquanto algoritmos de Aprendizagem Profunda necessitam de milhares de imagens para terem resultados significativos.

[0094] A Figura 12 representa de forma geral as etapas presentes em processos que utilizam a Aprendizagem Profunda, compreendendo a entrada, extração de características com treinamento de modelo e saída.

[0095] Baseado nos estudos divulgados pelo estado da técnica e nos dispositivos relacionados com aplicação de produtos na pele, foram feitos os testes das linhas de pesquisa que determinam o processo de detecção e reconhecimento de imagens labiais precisas, por meio do uso de inteligência artificial de acordo com a presente invenção.

[0096] A utilização de câmeras de infravermelho para o reconhecimento facial é uma abordagem válida, uma vez que esta é mais robusta a mudanças de iluminação e não depende da cor. Parte-se do pressuposto que existe uma diferença de volume entre os lábios e o entorno da boca.

[0097] Algumas abordagens testadas atualmente utilizam um dispositivo de infravermelho disponível, por exemplo Leap Motion, também sendo conhecido uma abordagem de câmeras infravermelho utilizando o Kinect da Microsoft.

[0098] O Leap Motion é um dispositivo que consiste em um sensor infravermelho desenvolvido e calibrado para captar e identificar os movimentos das mãos permitindo a interação humano-máquina através de gestos. Desta forma, quando o equipamento detecta a mão, a intensidade da luz infravermelha é ajustada e é possível visualizar claramente os lábios,

conforme visto na Figura 13A. Em comparação, em uma situação em que o equipamento não consegue detectar a mão, a intensidade da luz infravermelha é mantida na intensidade máxima e isso torna difícil identificar os lábios, como pode ser visto na Figura 13B e na Figura 13C.

[0099] O fato de ser possível visualizar os lábios em uma imagem infravermelha, permite considerar, que a abordagem pode auxiliar na detecção dos lábios bem como uma possível definição de borda dos lábios.

[00100] Existem diferentes abordagens para a detecção e segmentação dos lábios em imagens RGB. Essas abordagens podem se basear: em padrões de textura (uma vez que a textura da boca difere da textura ao redor); predição de pontos chave (Key Points) que representam as principais características do contorno dos lábios; utilização de reconhecimento de padrões, onde um modelo é treinado para identificar e segmentar a parte de interesse da imagem; segmentação por cor, uma vez que existe uma diferença na tonalidade entre a pele dos lábios e do entorno. Na utilização de pontos chave para representar os principais pontos/características do rosto, alguns pontos não são precisos em relação às bordas dos lábios, porém, isso é algo ajustável com o objetivo de se obter maior precisão e também compensado pelo uso de outras técnicas.

[00101] Imagens de alta resolução são imagens com alta quantidade de pixels, ou seja, permitem ver detalhes mínimos das imagens sem perda de qualidade mesmo com aumento do nível de zoom da imagem. Estas imagens permitem a visualização mais precisa de características da pele e da mucosa, tais como: linhas de expressão, digitais e relevos.

[00102] Atualmente, a aquisição de imagens de alta resolução é cada vez mais comum, uma vez que as câmeras capazes de adquirir esse tipo de imagem estão se popularizando e sendo encontradas em celulares, webcams e câmeras fotográficas.

[00103] A câmera Logitech C920 é um exemplo de webcam capaz de capturar imagens em full HD. Uma das vantagens desta câmera é a quantidade de configurações via software disponíveis como: foco, brilho, modo noturno, entre outras. A Figura 14 apresenta algumas imagens capturadas com este modelo de câmera, modificada para ter a menor configuração de foco possível (cerca de 2 cm da lente).

[00104] A utilização de imagens de microscopia simples (Figuras 15A a 15D) permite a visualização de características mais precisas da pele e da mucosa labial tais como linhas de expressão, digitais e tubérculos. Para a captura de imagens com alta resolução é necessário a utilização de lentes que permitam a aproximação desejada em relação ao objeto a ser capturado por uma câmera (podendo ser de captura RGB ou não). As Figuras 15A, 15B, 15C e 15D mostram exemplos de imagem de alta resolução, capturadas utilizando uma lente especial. A Figura 15A trata-se de uma foto da pele ao redor dos lábios e a Figura 15B uma imagem da região dos lábios. Ambas as imagens possuem um aumento de 60x. Por sua vez, a Figura 15C e 15D são, respectivamente, fotos da pele ao redor dos lábios e do lábio, ambas com 100x de aumento.

[00105] Normalmente ao passar produtos labiais os usuários podem definir o limite do produto como sendo apenas na área dos lábios ou passando para a região da pele ao redor. Uma boa definição das curvas e do desenho do pigmento na pele,

parece ser mais importante para os usuários do que a fixação do produto apenas na área dos lábios, conforme visto na Figura 16, que apresenta imagens de alta definição das curvas do desenho do lábio na aplicação do batom.

[00106] No desenvolvimento da presente invenção algumas considerações anteriores são aprofundadas e detalhadas. Outras abordagens são apresentadas de uma forma mais aprofundada. Para o aprofundamento dessas abordagens são realizados testes detalhados e elaboradas provas de conceito, que avaliam a viabilidade da abordagem.

[00107] A abordagem utilizando imagens em infravermelho levantada anteriormente consiste na captura de imagens em uma frequência de luz não visível (infravermelha), os benefícios desta abordagem envolvem principalmente a não interferência a luz visível, sendo assim mais fácil de se identificar um objeto no escuro ou sobre luz direta.

[00108] Como mencionado anteriormente, o Leap Motion é um dispositivo desenvolvido para o rastreo das mãos e identificação de gestos, utilizando sensores de infravermelho. Este equipamento tem um desempenho muito bom para o rastreo das mãos, contudo algumas limitações em seu programa computacional embarcado não permitem a utilização dele como um dispositivo de captura de imagens infravermelhas, estas limitações englobam o acesso às imagens capturadas, o controle de luminosidade dos LEDs infravermelhos e a compatibilidade com linguagem e sistema operacional.

[00109] Para determinação do processo da presente invenção são consideradas bases de imagens em infravermelho, incluindo bases com imagens da face em diferentes distâncias e

posições, e bases com vídeos em infravermelho e visão noturna, sendo escolhida ao menos uma base para os testes dos algoritmos e validação de aplicação. Dentre as bases encontradas para testes e resultados apresentados a seguir, é escolhida, por exemplo, a base SCface -Surveillance Cameras Face Database, sendo esta base composta por imagens da face em vários ângulos e capturadas por câmeras distintas, uma RGB e outra infravermelha. Um exemplo das imagens dessa base é apresentado na Figura 17.

[00110] No que se refere ao reconhecimento de padrões em imagens em infravermelho, pode ser feito utilizando diversas abordagens (modelos e técnicas) que podem ser aplicadas separadas ou combinadas. Entre essas abordagens, é escolhida a utilização de um modelo de predição de contornos, tal como a U-Net, divulgada pela publicação de Long, J. et al (Long, J., Shelhamer, E., Darrell, T. (2014). "Fully convolutional networks for semantic segmentation. In Computer Vision and Pattern Recognition). A U-Net é uma rede convolucional (CNN) desenvolvida para segmentação de imagens biomédicas e é principalmente utilizada para a segmentação de células em lâminas microscópicas, permitindo por exemplo a contagem das células. Esta rede é denominada fully convolutional e sua arquitetura é representada em um formato de "U".

[00111] Desta forma a rede recebe como entrada durante o processo de treinamento a imagem original em escala de cinza e uma máscara, ao longo do processamento da rede ela irá fazer o encoder e o decoder das imagens, gerando assim a máscara predita e ao fim do treinamento um modelo matemático capaz de prever máscaras do objeto usado no treinamento. Já durante o teste e uso do modelo gerado, apenas a imagem

original em escala de cinza é utilizada como entrada, e a saída do modelo é uma máscara de segmentação. As Figuras 18A e 18B apresentam um resultado original dessa rede, sendo que a imagem 18A representa a imagem de entrada para a predição e a imagem 18B apresenta os contornos de segmentação preditos pela rede.

[00112] Apesar da U-Net ser uma rede focada em imagens biomédicas, sua capacidade de identificar detalhes da imagem para utilizar na segmentação permite que essa rede seja utilizada em outras áreas. No contexto da presente invenção, considera-se que essa rede interpreta as informações de distância contidas nas imagens em infravermelho. Para fazer este teste, as etapas adotadas são apresentadas no diagrama da Figura 19, onde o dataset de imagens biomédicas (9) e suas respectivas máscaras (10), contendo 30 imagens, são utilizadas para o treinamento da rede (11) gerando assim um modelo de predição de máscaras (12).

[00113] Uma vez a rede treinada e um modelo preditivo criado, é utilizado um algoritmo haar cascade (14) para recortar as áreas de interesse da base de imagens em infravermelho (13), obtendo assim uma nova base (15) contendo, por exemplo 30 imagens para teste, sendo estas imagens são passadas para o modelo (16) que prediz as máscaras (17).

[00114] A Figura 20 apresenta alguns resultados de máscara preditos pela U-Net seguindo o processo e as observações apresentadas anteriormente, sendo na primeira coluna imagens de entrada para a rede (descrita no diagrama pela etapa (15)) e na segunda coluna a saída da rede (descrita no diagrama pela etapa (17)).

[00115] Ao observar os resultados podemos concluir que a rede consegue descrever uma máscara válida para algumas das imagens de entrada. É possível observar que nas imagens 2, 3, 6 e 7 os contornos são ressaltados, sendo que na imagem 6 existe uma predominância da barba na máscara e na imagem 7, que não contém a boca, mas apenas o queixo do usuário, a rede conseguiu delinear perfeitamente o contorno do rosto. Em contrapartida nas imagens 1, 4 e 5 a rede não consegue gerar uma máscara.

[00116] A utilização de imagens coloridas RGB para a detecção e segmentação de objetos é uma metodologia tendo diversos algoritmos, modelos matemáticos e técnicas de aprimoramento desenvolvidas com esse foco. Neste sentido são apresentadas algumas técnicas escolhidas, bem como os procedimentos de cada uma e seus resultados para a segmentação de boca e lábios.

[00117] Inicialmente, antes mesmo de realizar o treinamento de qualquer técnica de segmentação utilizando reconhecimento de padrões, é necessário rotular as imagens da base de treinamento, por exemplo, como apresentado na Figura 21. Esse processo tem a finalidade de, em linhas gerais, "mostrar" aos algoritmos de reconhecimento de padrões o que é lábio e que é isso que ele deve "aprender" para realizar a segmentação.

[00118] A rede U-Net descrita anteriormente para reconhecimento de padrões em imagens em infravermelho também pode ser utilizada para imagens RGB. Para isso existem duas abordagens iniciais. A primeira trata de uma modificação na arquitetura da rede, permitindo que ela reconheça imagens com três canais (RGB). Já a segunda consiste na conversão da imagem de entrada em uma imagem escala de cinza. A abordagem

escolhida é a de converter a imagem em escala de cinza, desta forma é possível obter os resultados dos testes de uma forma mais rápida.

[00119] Utilizando o dataset de imagens marcadas apresentado anteriormente sobre a marcação das imagens (ground truth) os testes realizados com a U-Net seguem a seguinte sequência de etapas:

- conversão das imagens de RGB para escala de cinza;
- conversão das marcações para máscaras de segmentação;
- treino da rede e criação do modelo e
- teste do modelo.

[00120] Uma vez que a U-Net é sensível à máscara de entrada durante o treinamento, diferentes formas de se criar as máscaras são testadas, sendo que para cada teste, a sequência descrita acima é refeita. A Figura 22 a seguir, mostra esse processo, sendo o passo (18) a imagem em escala de cinza, o passo (19) exemplos de diferentes versões de máscara criados, o passo (20) os modelos gerados para cada máscara, o passo (21) a imagem de entrada para o modelo predizer a máscara e o passo (22) as diferentes respostas de cada modelo para cada máscara.

[00121] Deste modo, para cada máscara é obtido um modelo distinto e assim uma predição de máscara diferente, as máscaras são variadas com relação à espessura da linha de marcação, inversão da área de interesse (representado nas máscaras pela região preta) e pela área de interesse. Exemplos destes diferentes testes podem ser observados na Figura 23 de resultados.

[00122] Como é possível observar na tabela de resultados, a máscara que permite uma melhor convergência do modelo, e

assim uma melhor predição da máscara, foi a máscara de borda de lábios com a linha de contorno mais espessa (line = 10) apresentada nas linhas 3 e 6 da Figura 23 de resultados.

[00123] A Figura 24 apresenta outros resultados para esse modelo, sendo as imagens da esquerda as originais em escala de cinza e as imagens da direita a predição feita pelo modelo.

[00124] Observando os resultados da Figura 24, pode-se verificar que mesmo a rede não trabalhando com imagens coloridas, ou seja sem a informação de cor, ela consegue mensurar o contorno da boca com certa precisão. Uma vez que identificamos que a informação de cor é extremamente relevante para a delimitação dos lábios, juntamente com as informações de textura usados por essa rede, a modificação da rede para utilizar as informações de cores das imagens RGB, juntamente com uma dataset maior, de acordo com a presente invenção, pode fazer com que a rede consiga ter um resultado melhor, permitindo que seu resultado seja usado por outras redes ou técnicas para a delimitação do contorno.

[00125] Como já mencionado anteriormente sobre a detecção, segmentação e reconhecimento de objetos em imagens digitais, há diferenças entre o processo de detecção de objetos em uma imagem e a segmentação de objetos em uma imagem. O processo de detecção de objetos em uma imagem consiste em fornecer as coordenadas dos Bounding Boxes (retângulos coloridos da Figura 25A) que são desenhados ao redor de cada objeto presente na imagem. Essas coordenadas são utilizadas para sinalizar a localização de cada objeto na imagem. Porém, esse processo de detecção de objetos não fornece nenhuma informação adicional além da localização dos objetos. Já o

processo de segmentação de objetos em uma imagem, inicialmente realiza o processo de detecção para localizar os objetos na imagem para depois segmentar os objetos. Na etapa de segmentação é criada uma máscara (Figura 25B) para cada objeto, a qual, fornece não só a localização dos objetos na imagem, como também informações de contorno, forma e também torna possível separar os objetos da imagem como todo, permitindo o processamento dos objetos de forma isolada.

[00126] Para o contexto do processo da presente invenção, inicialmente faz-se necessário identificar os lábios e separar estes do restante da pele do rosto. Como se trata de um processo de delimitação labial com etapas que permitam a detecção e a segmentação da região labial visando determinar e suportar um sistema inteligente de aplicação de produtos labiais nos lábios com alta precisão, não basta apenas detectar a boca, é necessário ter informações da forma e do contorno da boca, pois, isso será um dos fatores determinantes para a aplicação precisa do produto labial, tal como batom.

[00127] Tratando-se de alternativas em modos de execução, a técnica de reconhecimento de padrões se destaca entre as técnicas de detecção de objetos. A técnica de detecção de objetos Faster R-CNN, proposta no documento Ren, S., He et al (Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99) propicia bons resultados na detecção dos mais variados objetos.

[00128] No sentido de não apenas detectar os lábios, sendo preciso também segmentá-los, a arquitetura Mask R-CNN é uma

técnica bastante referenciada e utilizada na área de Reconhecimento de Padrões no que diz respeito à segmentação de objetos. O Mask R-CNN é um método de segmentação de objetos que consiste em uma extensão da técnica abordada no Faster R-CNN. O Mask R-CNN, inicialmente, necessita realizar a detecção dos objetos presentes na imagem para em seguida realizar a segmentação de tais objetos. Então, em um primeiro momento a arquitetura da rede do Faster R-CNN é utilizada para realizar a detecção dos objetos fornecendo as coordenadas e, conseqüentemente, a localização de cada objeto na imagem, e após isso com a informação da detecção, de forma paralela, é feita a segmentação dos objetos através da criação de máscaras.

[00129] Tanto o Faster R-CNN como o Mask R-CNN são abordagens utilizadas para realizar a detecção e segmentação, respectivamente, dos mais variados objetos. Alguns ajustes e adaptações são realizadas no Mask R-CNN para atender ao contexto da presente invenção.

[00130] Tais ajustes e adaptações consistem em treinar e testar os algoritmos do Mask R-CNN com imagens dos lábios ou partes dos lábios, conforme visto nas Figuras 26A à 26F, e não de vários objetos encontrados no dia a dia como o algoritmo é usualmente treinado e testado por padrão.

[00131] Alguns testes são realizados utilizando a Mask R-CNN com o objetivo de verificar a viabilidade de utilizar essa técnica para separar a região dos lábios do restante da pele do rosto e o quão precisa é a segmentação realizada por essa abordagem.

[00132] Os testes preliminares que serão mostrados a seguir objetivam treinar o algoritmo do Mask R-CNN com uma base de

imagens de treinamento dos lábios ou parte dos lábios. Após o treinamento, será gerado um Modelo de Segmentação, o qual será usado para realizar a segmentação da região dos lábios das imagens de teste contendo a região dos lábios ou parte dos lábios.

[00133] Com o processo de rotular as imagens da base de treinamento concluído, os testes são realizados. Três cenários para os testes são estabelecidos e são descritos a seguir considerando a utilização de imagens da boca completa e de partes da boca da base de treinamento.

[00134] Em um primeiro cenário, conforme visto no diagrama da Figura 27, o objetivo principal é verificar a viabilidade de se utilizar o Mask R-CNN para segmentar a região dos lábios e o quão precisa é a segmentação realizada pela técnica, visto que esta é utilizada para realizar a segmentação de objetos variados presentes no dia-a-dia e no caso da presente invenção há um tipo específico de segmentação que deverá ser realizada e com precisão, os lábios. Neste primeiro cenário são considerados os seguintes parâmetros:

a) base de imagens de treinamento:

- número de imagens: 110 imagens
- contexto das imagens: boca completa e sem batom

b) base de imagens de teste:

- número de imagens: 94 imagens
- contexto das imagens: boca completa e sem batom

[00135] Em um segundo cenário, conforme visto no diagrama da Figura 28, o objetivo é avaliar a capacidade do modelo de segmentação de aprender a diferenciar a pele dos lábios da pele do rosto sem possuir o contexto da boca por completo,

visto que há a possibilidade de ser necessário o dispositivo final estar muito próximo dos lábios não tendo o contexto completo da boca na aquisição das imagens no momento da aplicação do batom. Esse aprendizado ocorre em sua maior parte pelas características estruturais e de textura da pele dos lábios. Neste segundo cenário são considerados os seguintes parâmetros:

a) base de imagens de treinamento:

- número de imagens: 120 imagens
- contexto das imagens: parte dos lábios e sem batom

b) base de imagens de teste

- número de imagens: 92 imagens
- contexto das imagens: parte dos lábios

[00136] Em um terceiro cenário, conforme visto no diagrama da Figura 29, o objetivo principal é avaliar a capacidade de generalização do modelo de segmentação. O treinamento do algoritmo é realizado utilizando a base de imagens com partes dos lábios e o teste é feito utilizando imagens contendo a boca completa. Avalia-se a capacidade do modelo de segmentação em ter sido treinado com partes dos lábios e conseguir generalizar a segmentação para a boca completa e não apenas para partes dos lábios. Treinar o algoritmo com apenas partes dos lábios, apesar de não ter o contexto completo da boca, favorece a uma maior extração das características estruturais dos lábios devido ao zoom que é dado no momento da aquisição da imagem. Neste terceiro cenário são considerados os seguintes parâmetros:

a) base de imagens de treinamento:

- número de imagens: 120 imagens
- contexto das imagens: parte dos lábios e sem batom

b) base de imagens de teste:

- número de imagens: 94 imagens
- contexto das imagens: boca completa

[00137] A seguir são apresentados os resultados decorrentes dos treinamentos e testes citados nos cenários acima.

[00138] No primeiro cenário, no qual o algoritmo foi treinado com imagens da boca completa e testado com imagens também da boca completa, a acurácia obtida foi de 70%, ou seja, a região dos lábios foi segmentada corretamente em 70% das imagens de teste. Isso mostra a viabilidade de utilizar o Mask R-CNN como parte da solução da presente invenção. Conforme visto na Figura 30, são apresentados alguns dos resultados do processo de segmentação dos lábios considerados satisfatórios através da avaliação de um especialista.

[00139] No segundo cenário, no qual o algoritmo é treinado com imagens de partes dos lábios e testado com imagens também de partes dos lábios, a acurácia obtida foi de 47%, ou seja, a região dos lábios foi segmentada corretamente em 47% das imagens de teste. O objetivo deste teste é verificar se o modelo de segmentação conseguiria segmentar corretamente as regiões correspondentes aos lábios sem haver o contexto da boca completa na imagem, visto que o dispositivo final pode exigir uma distância muito próxima dos lábios para aquisição das imagens e não obter o contexto completo da boca. Através destes resultados observa-se que apesar de a acurácia ter sido menor, se comparada com a acurácia obtida no primeiro cenário, o modelo conseguiu segmentar as regiões referentes aos lábios em imagens sem o contexto da boca completa. Conforme visto na Figura 31, são apresentados alguns dos resultados considerados satisfatórios do processo de

segmentação dos lábios através da avaliação de um especialista.

[00140] No terceiro cenário, no qual o algoritmo foi treinado com imagens de partes dos lábios e testado com imagens da boca completa, a acurácia obtida é de 3%, ou seja, a região dos lábios foi segmentada corretamente em 3% das imagens de teste. O objetivo deste teste é verificar a capacidade do modelo de segmentação de generalizar a segmentação dos lábios. Já é esperada que haja uma queda na acurácia do modelo de segmentação, dado que o algoritmo foi treinado com imagens de partes dos lábios e no teste foi forçado a segmentar a região dos lábios com imagens contendo o contexto da boca completa, as quais, não foram apresentadas ao algoritmo durante o treinamento. O modelo de segmentação foi forçado a sair de um aprendizado com imagens com representações "micro" (menor contexto) dos lábios para realizar a segmentação dos lábios em imagens com uma representação dos lábios "macro" (maior contexto). Conforme visto na Figura 32 e na Figura 33, são apresentados alguns resultados do processo de segmentação dos lábios satisfatórios e insatisfatórios respectivamente.

[00141] No que se refere à técnica de superpixel, consiste no agrupamento dos pixels de uma imagem com base na similaridade da característica de cor. O método utiliza técnicas de processamento de imagens convencionais sem a necessidade de aprendizagem profunda, como é feito nas abordagens apresentadas anteriormente (deep learning). O conceito utilizado para a realização do agrupamento dos pixels é o algoritmo de clusterização k-means, que agrupa elementos em dado espaço de características similares, tendo

como cálculo base a distância euclidiana entre dois pontos. A técnica de clusterização é utilizada em machine learning para casos em que a aprendizagem é não-supervisionada, ou seja, casos em que os objetos não possuem classes predefinidas. O agrupamento dos objetos no espaço faz com que a determinação do aprendizado seja apontada pelo grupo ao qual ele pertence.

[00142] Para melhor exemplificar tal comportamento, a Figura 34 ilustra um espaço de cor de uma imagem digital (RGB) e o agrupamento realizado. É possível perceber que a distribuição das cores dos pixels de uma imagem possui um padrão que pode vir representar objetos na imagem.

[00143] Tal fato pode ser também observado na própria imagem digital, em que o método de superpixel torna possível distinguir os objetos que ela possui. Por exemplo, a imagem da Figura 35A trata-se da imagem real, e a imagem da Figura 35B trata-se da imagem clusterizada. Os elementos que compõem a imagem são segmentados no sentido de que áreas de cores muito próximas são consideradas objetos de uma mesma classe, um exemplo são os objetos relacionados às árvores e às folhagens no plano de fundo que foram classificadas como zona verde na imagem, tal fato serve para os demais elementos como região do campo de rosas e o céu.

[00144] Para análise da segmentação dos elementos nas imagens que tratam-se de pele da face e pele da boca, é utilizado um algoritmo slic, tal como proposto por Achanta, R. et al (Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2010). Slic superpixels (No. REP_WORK)), que realiza a clusterização (agrupamentos) com o método k-means. Os principais parâmetros para a função slic são o número de segmentos, ou seja, o número de clusters que deseja

separar dos elementos e a imagem a ser processada. Neste contexto, o algoritmo foi aplicado às imagens recortadas do dataset de lábios, conforme visto nas Figuras 36A à 36E (borda de lábio sem pigmento), 37A à 37E (borda de lábio com pigmento), e 38A à 38E (borda de lábio sem pigmento), o número de segmentos utilizado foi de 30, 20, 10 e 2 (foi considerado esse valor para verificar se o algoritmo delimita a zona labial e a zona da pele).

[00145] A fim de obter um melhor resultado para a segmentação dos lábios com o número de segmentos definido como 2, algumas transformações acerca do espaço de cor são realizadas, como a mudança da matiz, saturação e valor (espaço HSV - Hue, Saturation, Value), a fim de tornar os elementos da imagem perceptíveis ao algoritmo de superpixel, em que não houveram separações bem sucedidas. Tal comportamento é apresentado nas imagens das Figuras 39A à 39C, que apresentam imagens de zoneamento por superpixel aplicado em borda de lábio sem pigmento no espaço de cor HSV. A Figura 39A representa uma imagem real, a 39B representa uma imagem em HSV e a Figura 39C representa uma imagem em HSV com superpixel de 2 segmentos.

[00146] Importante notar que no espaço de cor HSV há uma melhora na separação devido ao contraste que a saturação proporcionou à nova imagem por causa da iluminação incidente no lábio, entretanto em alguns casos o espaço de cor RGB mostra melhor desempenho, conforme visto nas Figuras 40A à 40C. A Figura 40A representa uma imagem real, a 40B representa uma imagem real com a máscara de processamento HSV e a Figura 40C representa uma imagem real com a máscara de processamento RGB.

[00147] No sentido de melhorar a precisão da segmentação dos lábios, a presente invenção une as duas abordagens, antes testadas separadamente: o Superpixel e a Mask R-CNN.

[00148] A junção inusitada dos métodos de Superpixel e Mask R-CNN de acordo com a presente invenção objetiva utilizar as informações de separação entre os lábios e pele do rosto resultantes da execução da abordagem do Superpixel no treinamento do algoritmo da abordagem da Mask R-CNN. Tais informações de separação dos superpixels são inseridas nas imagens utilizadas no treinamento da Mask R-CNN de acordo com o fluxo de execução apresentado na Figura 41.

[00149] Este fluxo é executado em todas as imagens de partes dos lábios da base de treinamento e todas essas imagens são utilizadas no fluxo de execução inferior (Mask R-CNN - Treinamento) para o treinamento do algoritmo. As imagens de partes dos lábios da base de treinamento foram utilizadas ao invés das imagens da boca completa por possuírem um caráter mais desafiador para o modelo de segmentação, pois não possuem o contexto da boca completa e possuem uma quantidade menor de características a serem aprendidas pelo algoritmo se comparadas com as imagens da boca completa.

[00150] Na Figura 41, em uma primeira etapa de pré-processamento utilizando Superpixel, uma imagem original (23) de entrada do fluxo de execução é submetida à etapa de abordagem da segmentação da imagem por Superpixel com a extração de contornos (24) resultando na imagem com a separação entre lábio e pele do rosto (25). Ainda nesta primeira etapa de pré-processamento por superpixel ocorre a etapa de extração da máscara (26) referente à imagem com a

separação entre lábio e pele do rosto (25) e as informações dessa máscara são inseridas na imagem original (23).

[00151] Ainda na etapa de extração da máscara (26), para inserir as informações da máscara na imagem de entrada, primeiro converte-se o espaço de cor da imagem original (23) de RGB para HSV.

[00152] Em seguida, insere-se as informações da máscara no canal de luminância V com o objetivo de exaltar essa separação entre lábio e pele do rosto na imagem RGB final. Após realizar a inserção das informações da máscara no canal de luminância V, a imagem no espaço de cor HSV é convertida para o espaço de cor RGB, resultando na imagem resultante (27) a ser inserida no processo de treinamento utilizando o algoritmo da Mask R-CNN. Com a base de imagens de parte dos lábios de treinamento resultante da etapa de pré-processamento utilizando Superpixel, a etapa de treinamento utilizando o algoritmo da Mask R-CNN é continuada pela etapa de execução do treinamento da técnica de segmentação baseada no reconhecimento de padrões Mask R-CNN (28) e por fim é realizada a obtenção do modelo de segmentação (29).

[00153] Para os testes realizados seguindo essa estratégia, há dois cenários, sendo que em um primeiro cenário, o objetivo principal é verificar se alguma melhoria significativa é obtida utilizando imagens de partes dos lábios tanto na base de treinamento como na base de testes. A acurácia obtida foi de 23%, ou seja, a região dos lábios foi segmentada corretamente em 23% das imagens de teste, conforme apresentado nas imagens da Figura 42.

[00154] Em um segundo cenário, o objetivo principal é verificar a capacidade do modelo de segmentação de

generalizar o conhecimento obtido baseado em uma base de imagens de treinamento de partes dos lábios para imagens de teste com o contexto da boca completa. A acurácia obtida foi de 7%, ou seja, a região dos lábios foi segmentada corretamente em 7% das imagens de teste, conforme apresentado nas imagens da Figura 43.

[00155] As diversas etapas dos métodos de processamento aqui apresentados são realizadas por meio de um módulo de processamento presente em um sistema de aplicação de produtos labiais. Neste módulo de processamento é realizado todo o processamento necessário para a identificação das bordas labiais, após a captura das informações. Na etapa de predição do modelo é o momento em que módulo de processamento define as regiões labiais que devem receber o pigmento pelo sistema de aplicação de produtos labiais. Essas regiões labiais definidas pelo modelo passam por uma etapa de pós-processamento necessária para garantir o formato correto das informações que sejam compatíveis com o sistema de aplicação de produtos labiais. Em seguida, o sistema de aplicação utiliza o pigmento e aplica na região labial identificada anteriormente.

REIVINDICAÇÕES

1. "MÉTODO DE DETECÇÃO E SEGMENTAÇÃO DA REGIÃO LABIAL", caracterizado por compreender as etapas de:

- realizar o reconhecimento de padrões por meio da extração de características de entrada de imagens labiais, rotulando-as para uma base de treinamento por meio de um módulo de processamento presente em um sistema de aplicação de produtos labiais;

- definir e indicar as imagens labiais rotuladas aos algoritmos de reconhecimento de padrões para aprendizagem e realização de segmentação das ditas imagens labiais pelo dito módulo de processamento; e

- executar o treinamento de um modelo em aprendizagem de máquina no dito módulo de processamento com uma pluralidade de dados exemplificados e suas respectivas respostas definindo rótulos que o modelo deve aprender e predizer para identificar e gerar um padrão matemático para um sistema de aplicação de produtos labiais.

2. "MÉTODO DE DETECÇÃO E SEGMENTAÇÃO DA REGIÃO LABIAL", de acordo com a reivindicação 1, caracterizado por:

- a etapa de realizar o reconhecimento de padrões por meio da extração de características de entrada de imagens labiais compreende executar o reconhecimento de padrões em imagens em infravermelho, utilizando um modelo de predição de contornos por uma rede convolucional (CNN) U-Net; e

- a etapa de executar o treinamento de um modelo em aprendizagem de máquina no dito módulo de processamento compreende executar a codificação e decodificação de imagens labiais originais recebidas em escala de cinza e uma máscara como entrada na rede convolucional (CNN) U-Net durante o

processo de treinamento, gerar uma máscara predita, e ao fim do treinamento gerar um modelo matemático de predição de máscaras das imagens labiais utilizadas no treinamento.

3. "MÉTODO DE DETECÇÃO E SEGMENTAÇÃO DA REGIÃO LABIAL", de acordo com a reivindicação 1, caracterizado pela etapa de realizar o treinamento de um modelo em aprendizagem de máquina no dito módulo de processamento compreender:

- treinar um algoritmo do Mask R-CNN com uma base de imagens de treinamento dos lábios ou parte dos lábios de modo a aprender a diferenciar a pele dos lábios da pele do rosto; e

- gerar um modelo de segmentação da região das imagens contendo a região dos lábios ou parte dos lábios.

4. "MÉTODO DE DETECÇÃO E SEGMENTAÇÃO DA REGIÃO LABIAL", de acordo com a reivindicação 1, caracterizado por a etapa de realizar o treinamento de um modelo em aprendizagem de máquina no dito módulo de processamento compreender:

- executar um agrupamento dos pixels de uma imagem com base na similaridade da característica de cor por meio de um algoritmo de clusterização que agrupa elementos em dado espaço de características similares de modo que a determinação do aprendizado seja apontada pelo grupo ao qual ele pertence;

- gerar um modelo de segmentação da região das imagens contendo a região dos lábios ou parte dos lábios utilizado um algoritmo slic que realiza a clusterização com o método k-means utilizando parâmetros de número de segmentos, sendo o algoritmo aplicado à imagens recortadas da base de dados de imagens contendo a região dos lábios ou parte dos lábios; e

- executar a mudança de matiz, saturação e valor no

espaço de cor HSV, ou o espaço de cor RGB, de modo a tornar os elementos da imagem perceptíveis ao algoritmo de superpixel.

5. "MÉTODO DE DETECÇÃO E SEGMENTAÇÃO DA REGIÃO LABIAL", de acordo com a reivindicação 1, caracterizado por:

- em uma etapa de pré-processamento, submeter uma imagem labial original de entrada à etapa de segmentação da imagem por superpixel com a extração de contornos resultando na imagem com a separação entre lábio e pele do rosto;

- extrair uma máscara referente à imagem com a separação entre lábio e pele do rosto, inserir as informações dessa máscara na imagem original, e converter o espaço de cor da imagem original de entrada de RGB para HSV;

- inserir as informações da máscara no canal de luminância V de modo a exaltar a separação entre lábio e pele do rosto na imagem RGB final;

- converter a imagem no espaço de cor HSV para o espaço de cor RGB, obtendo uma imagem resultante;

- inserir a imagem resultante no processo de treinamento utilizando o algoritmo da Mask R-CNN; e

- executar o treinamento de segmentação utilizando o algoritmo da Mask R-CNN com a base de imagens de parte dos lábios de treinamento resultante da etapa de pré-processamento; e

- gerar um modelo de segmentação.

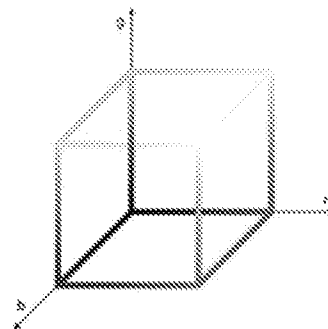
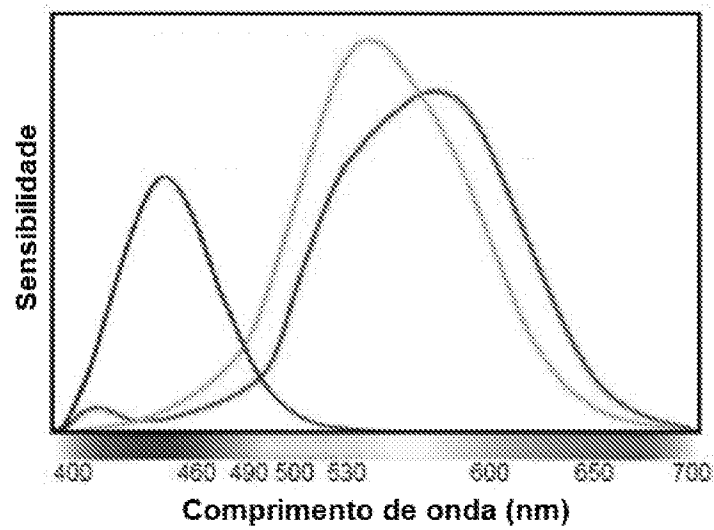
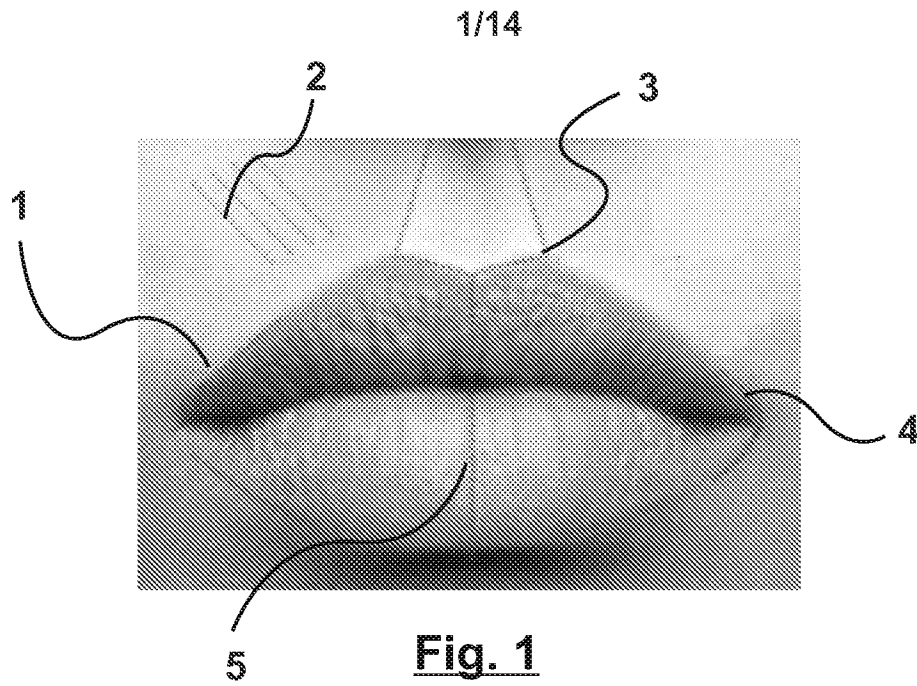


Fig. 3

2/14

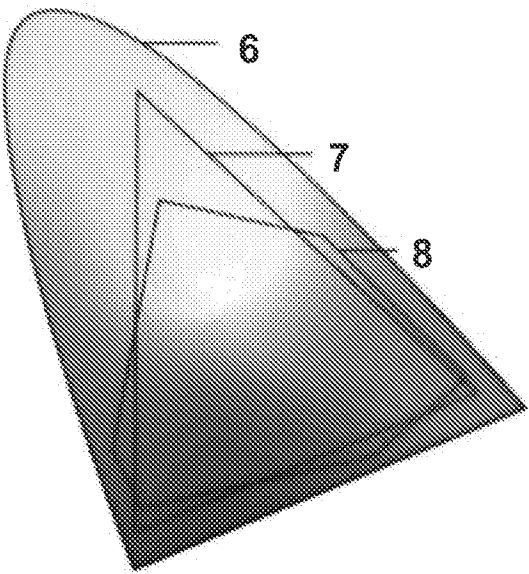


Fig. 4

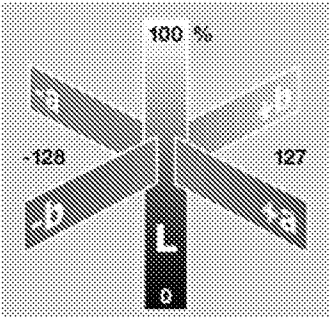


Fig. 5

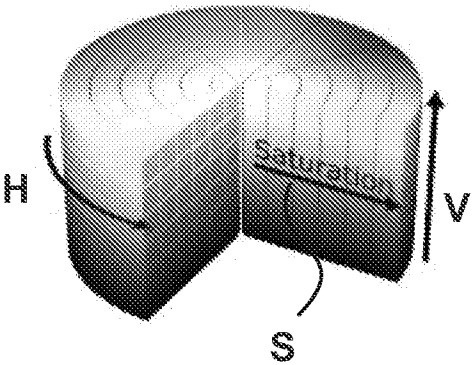


Fig. 6

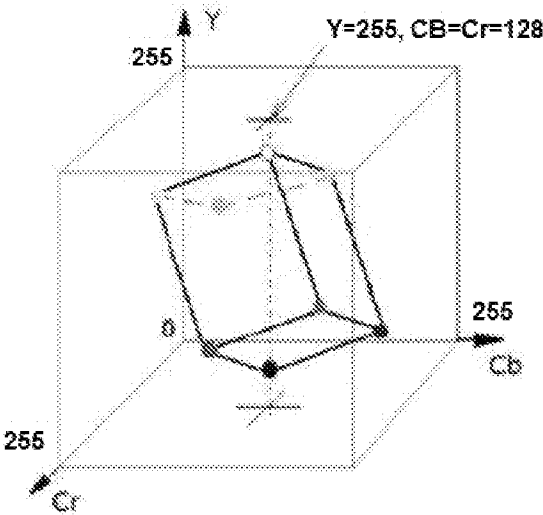


Fig. 7

3/14



Fig. 8

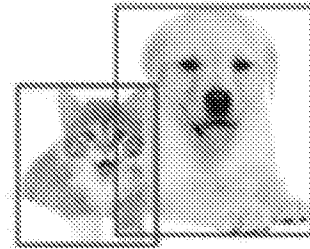


Fig. 9



Fig. 10

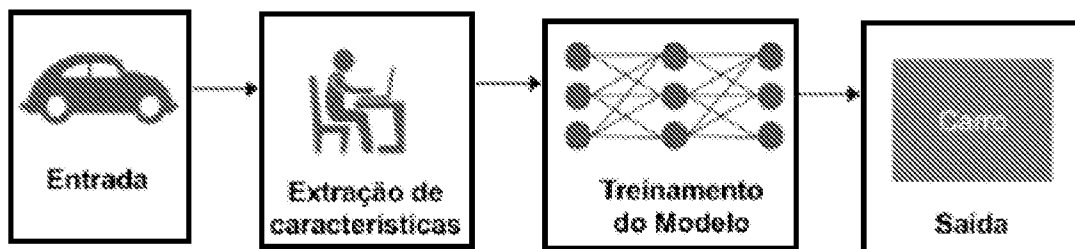


Fig. 11

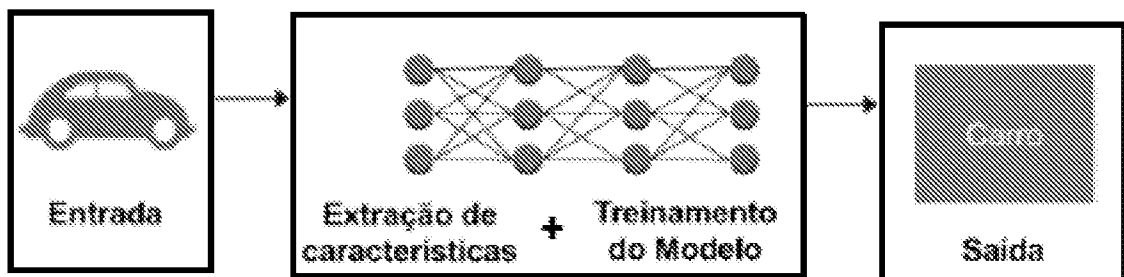


Fig. 12

4/14

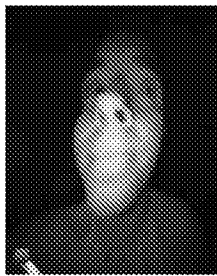


Fig. 13A

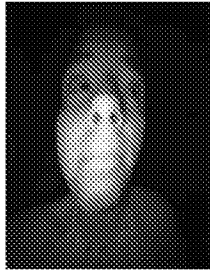


Fig. 13B

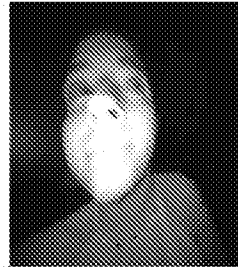


Fig. 13C

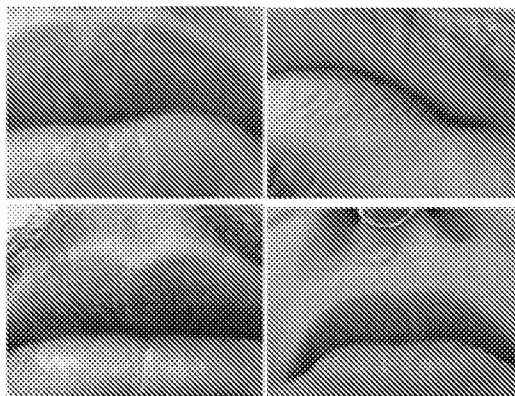


Fig. 14

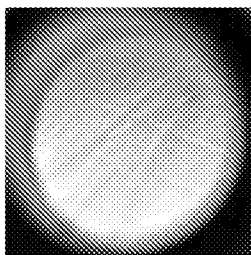


Fig. 15A

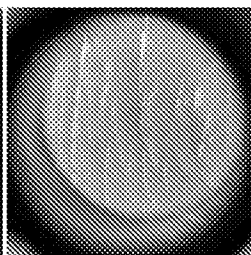


Fig. 15B

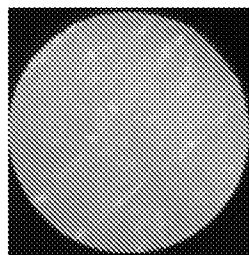


Fig. 15C

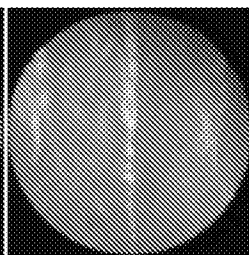


Fig. 15D

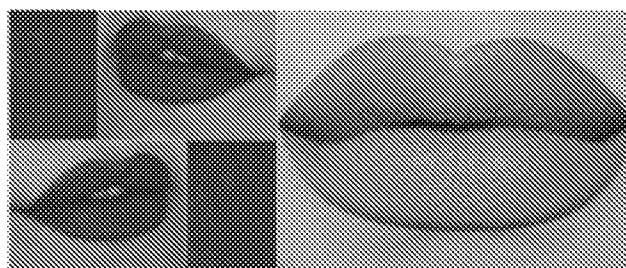


Fig. 16

5/14

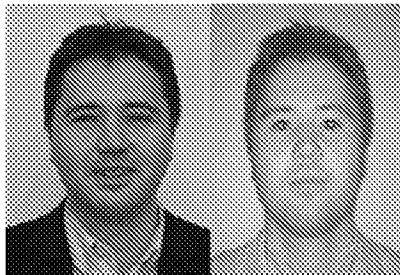


Fig. 17

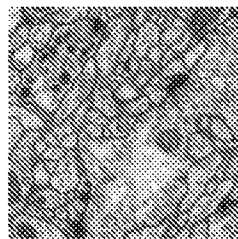


Fig. 18A



Fig. 18B

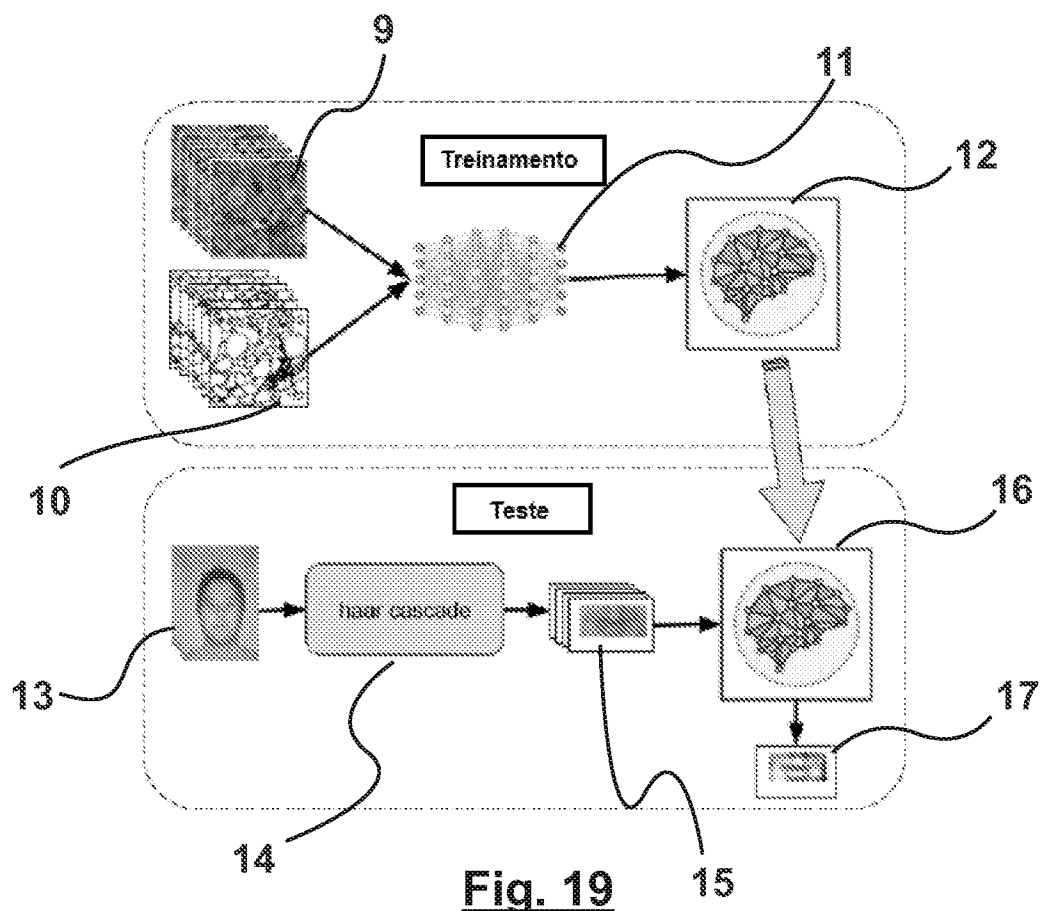


Fig. 19

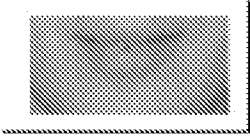
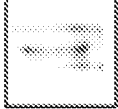
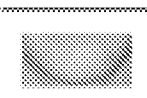
	Entrada	Saída
1		
2		
3		
4		
5		
6		
7		

Fig. 20

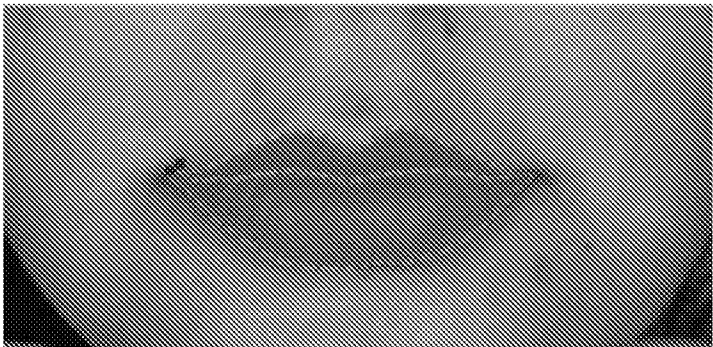


Fig. 21

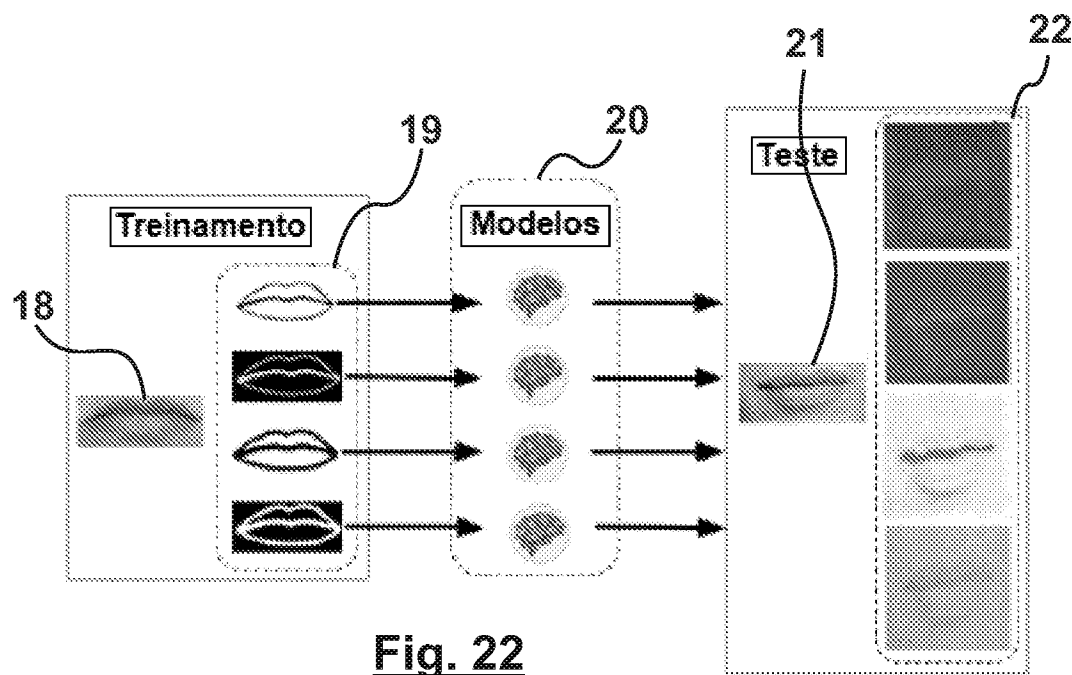


Fig. 22

	Imagem original {Treinamento}	Máscara {Treinamento}	Imagem original {Teste}	Máscara predita
1				
2				
3				
4				
5				
6				

Fig. 23

8/14

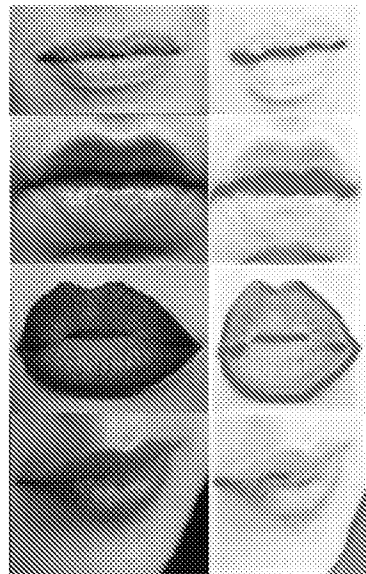


Fig. 24

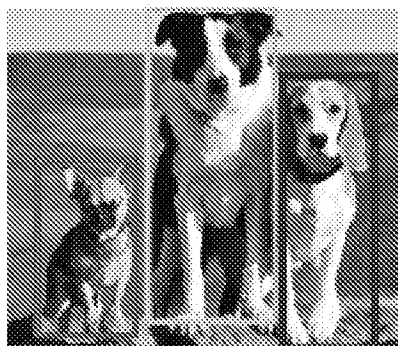


Fig. 25A

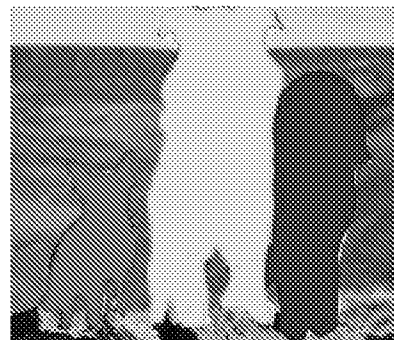


Fig. 25B

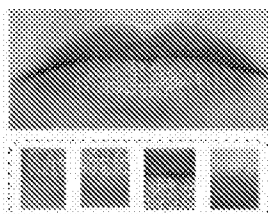


Fig. 26A

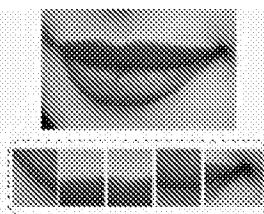


Fig. 26B

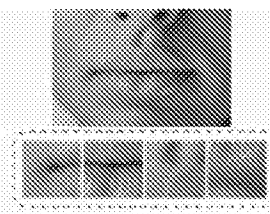


Fig. 26C

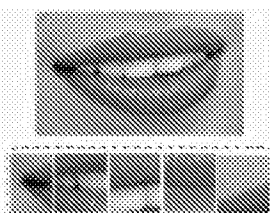


Fig. 26D

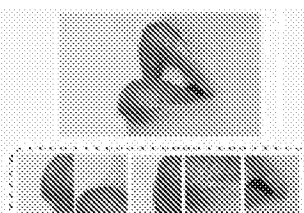


Fig. 26E

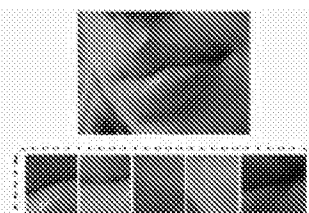
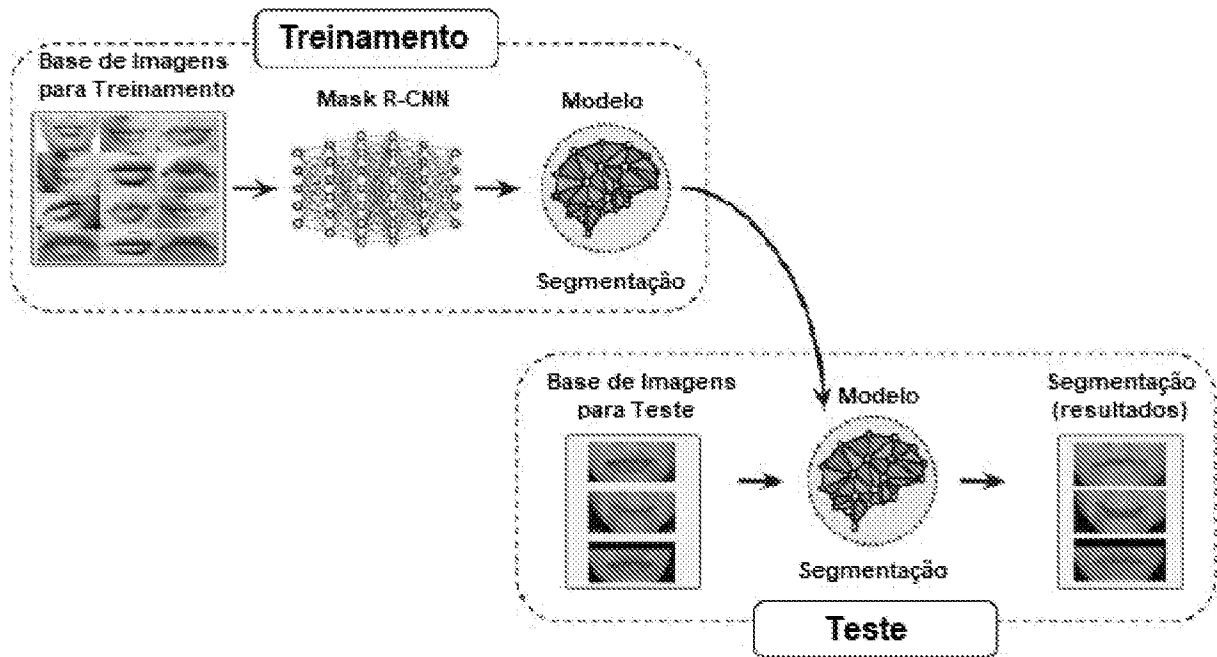
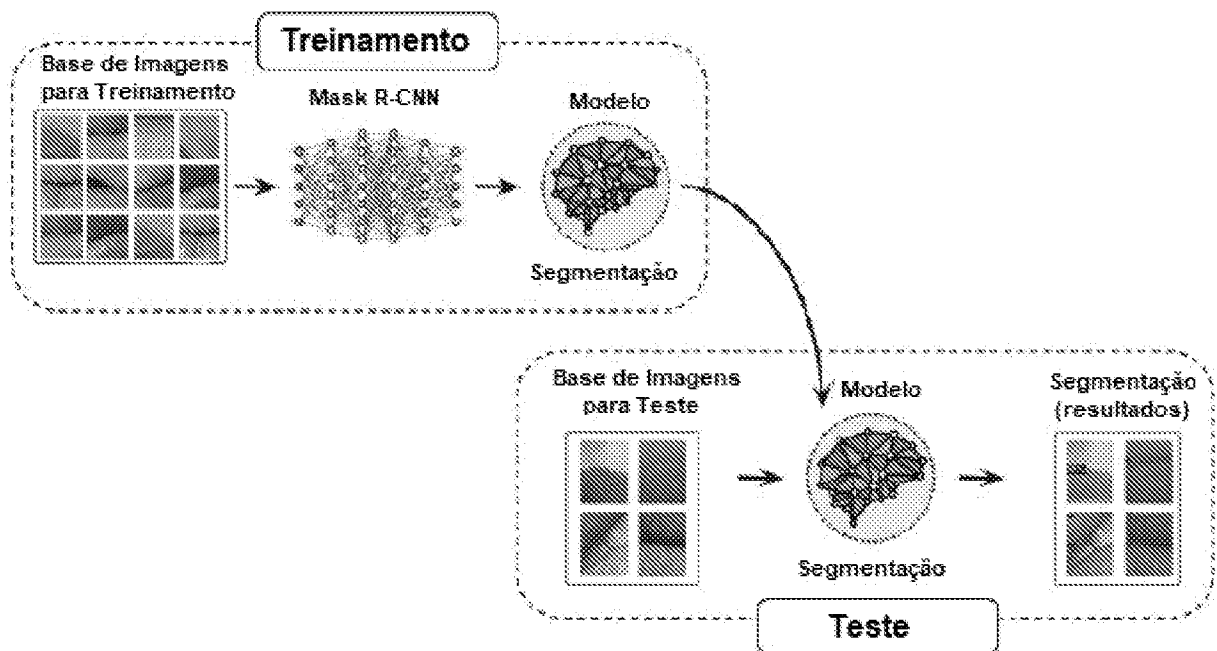
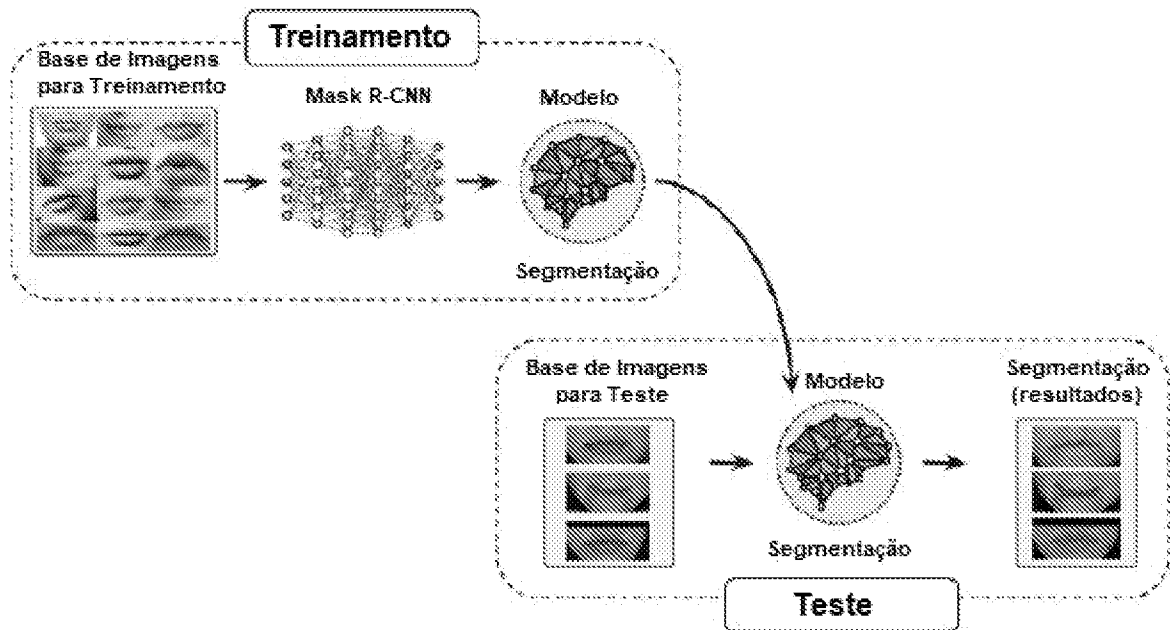
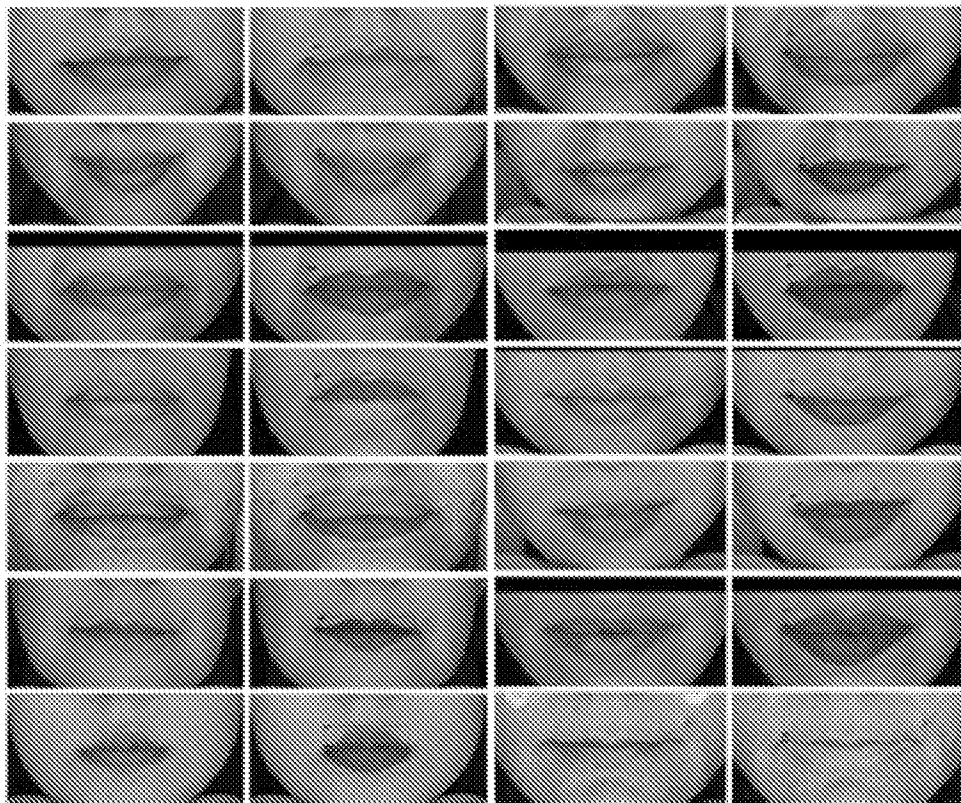


Fig. 26F

9/14

**Fig. 27****Fig. 28**

10/14

**Fig. 29****Fig. 30**

11/14

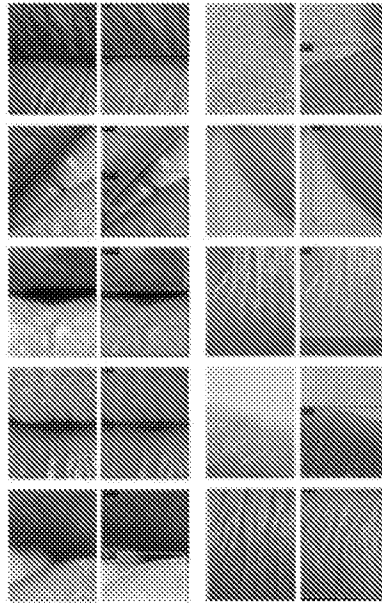


Fig. 31

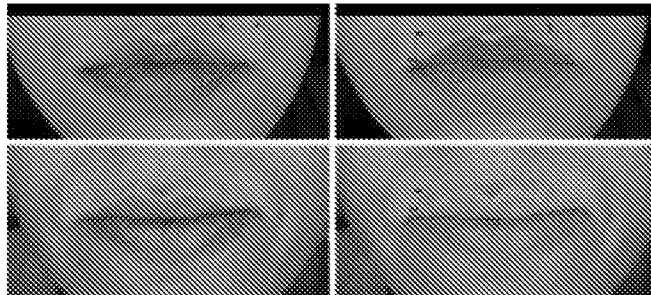


Fig. 32

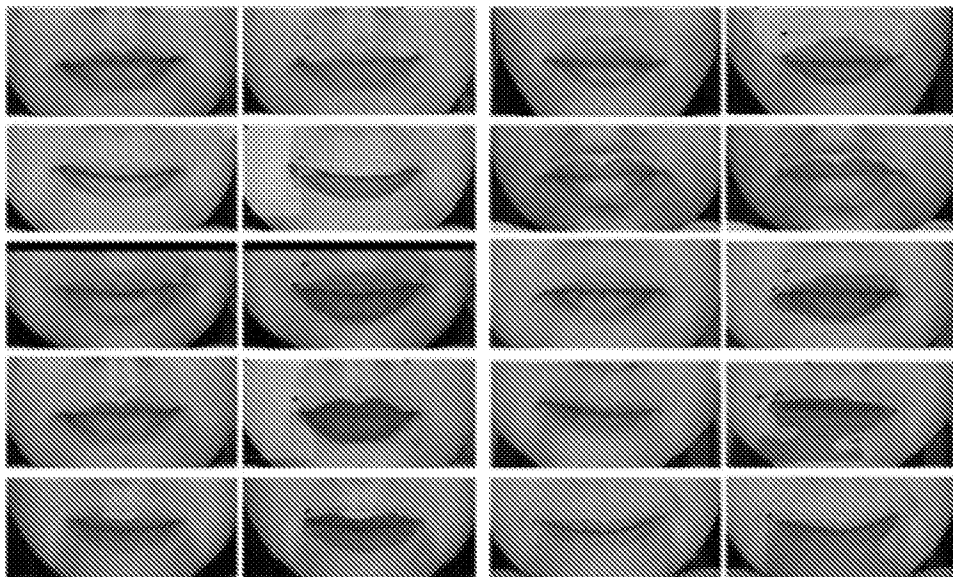


Fig. 33

12/14

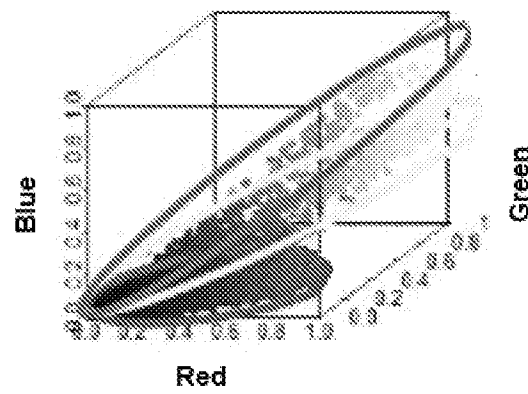


Fig. 34

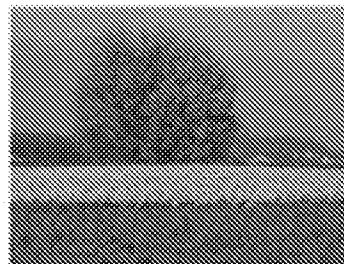


Fig. 35A

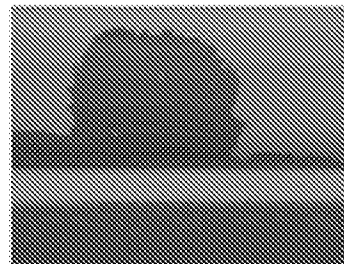


Fig. 35B

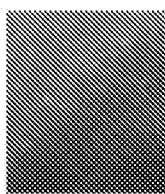


Fig. 36A

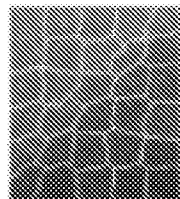


Fig. 36B

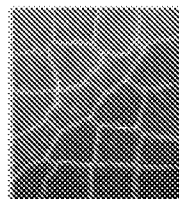


Fig. 36C

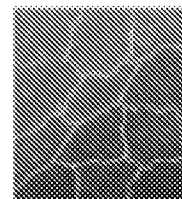


Fig. 36D

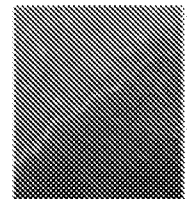


Fig. 36E

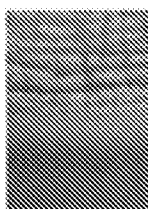


Fig. 37A

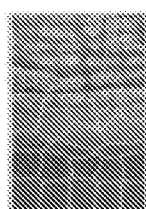


Fig. 37B

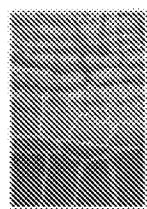


Fig. 37C

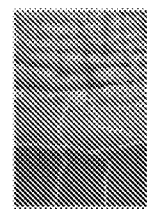


Fig. 37D

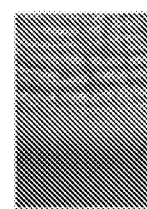
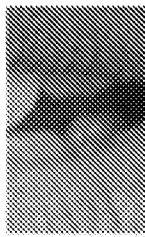
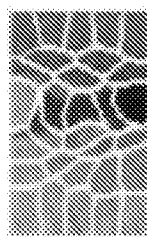
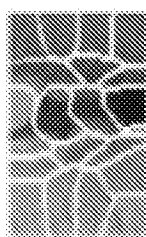
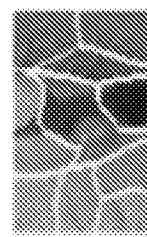
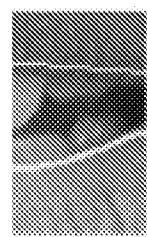
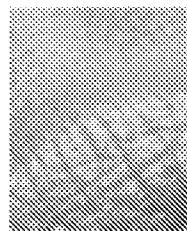
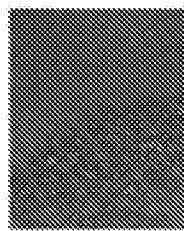
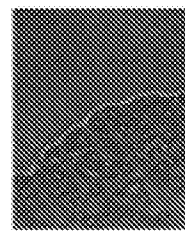
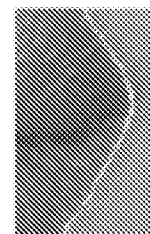
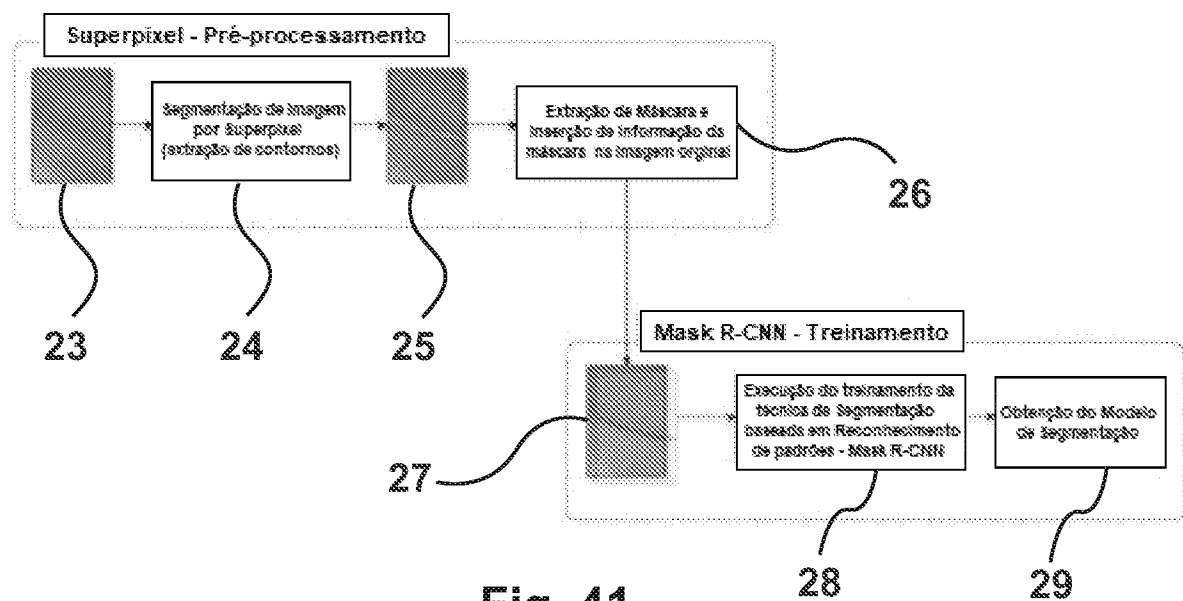


Fig. 37E

13/14

Fig. 38AFig. 38BFig. 38CFig. 38DFig. 38EFig. 39AFig. 39BFig. 39CFig. 40AFig. 40BFig. 40CFig. 41

14/14

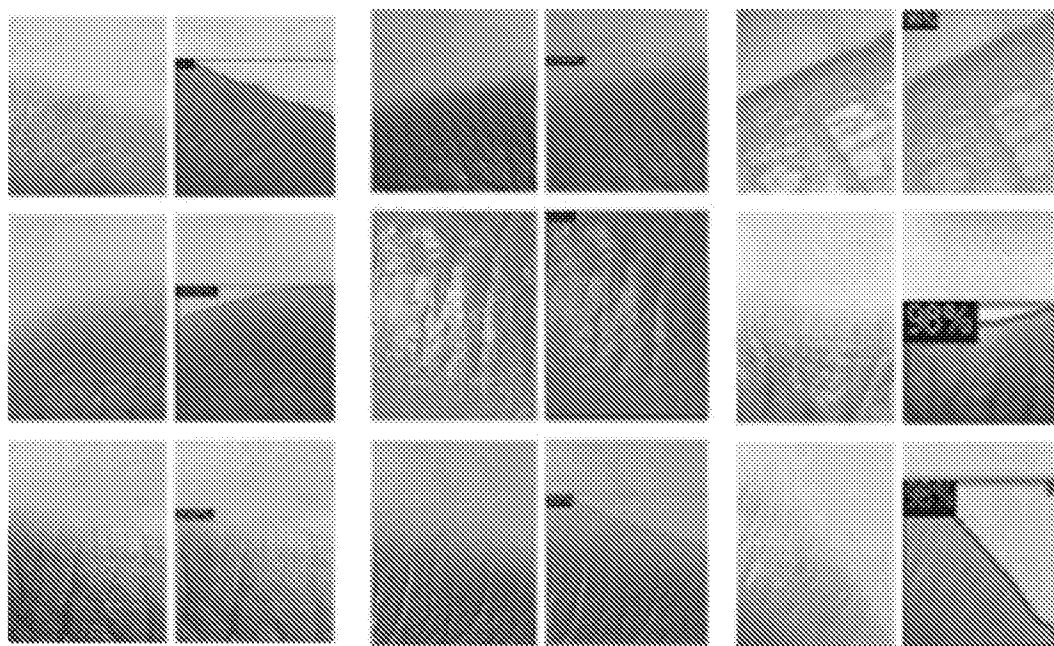


Fig. 42



Fig. 43

INTERNATIONAL SEARCH REPORT

International application No.

PCT/BR2021/050478

A. CLASSIFICATION OF SUBJECT MATTER

IPC: G06T7/11 (2017.01), G06K9/62 (2006.01), G06V10/25 (2022.01)

CPC: G06T7/11, G06K9/6271, G06V10/25

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06T, G06K, G06V

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Base de Patentes INPI-BR

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DERWENT, GOOGLE PATENTS, GOOGLE SCHOLAR

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	AHMAD, S., and FULLER, B. "Unconstrained iris segmentation using convolutional neural networks." In Asian Conference on Computer Vision, pp. 450-466. Springer, Cham, 2018.	1, 3
A	Obtained in: https://arxiv.org/pdf/1812.08245.pdf Access in 22/02/2022 (abstract; Sections 1, 2, 4.1)	2, 4, 5
X	RONNEBERGER, O., FISCHER, P., BROX, T., "U-Net:	1, 2
A	Convolutional Networks for Biomedical Image Segmentation In: NAVAB, N., HORNEGGER, J., WELLS, W., FRANGI, A. (eds) Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham. Obtained in: https://doi.org/10.1007/978-3-319-24574-4_28 Access in 22/02/2022 (abstract; Sections 1, 3)	3 – 5

☒ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

22 February 2022 (22.02.2022)

Date of mailing of the international search report

28 February 2022 (28.02.2022)

Name and mailing address of the ISA/ BR

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

 International application No.
PCT/BR2021/050478

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HE, K., Gkioxari, G., Dollár, R. and Girshick, R., "Maskr-cnn." In: Proceedings of the IEEE International conference on Computer vision, pp. 2961-2969. 2017.	1, 3
A	Obtained in: https://openaccess.thecvf.com/content_ICCV_2017/papers/He_Mask_R.-CNN_ICCV_2017_paper.pdf Access in 22/02/2022 (abstract; Sections 1 - 3, 5)	2, 4, 5
X	US 10713794 B1 (FACEBOOK INC [US]) 14 July 2020 (2020-07-14)	1, 3
A	(abstract, Figs. 9 - 14A; col. 35, 1. 43 - col. 38,	2, 4, 5
X	ACHANTA, R., SHAJI, A., SMITH, K., LUCCHI, A., FUA, P. and SLISSTRUNK, S., "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 11, pp. 2274-2282, Nov. 2012, doi: 10.1109/TPAMI.2012.120.	4
A	Obtained in: https://ieeexplore.ieee.org/document/6205760 Access in 22/02/2022 (abstract; p. 2275 - 2277, section 3)	1 - 3, 5

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/BR2021/050478

US 10713794 B1

2020-07-14

None

A. CLASSIFICAÇÃO DO OBJETO IPC: G06T7/11 (2017.01), G06K9/62 (2006.01), G06V10/25 (2022.01) CPC: G06T7/11, G06K9/6271, G06V10/25 De acordo com a Classificação Internacional de Patentes (IPC) ou conforme a classificação nacional e IPC		
B. DOMÍNIOS ABRANGIDOS PELA PESQUISA Documentação mínima pesquisada (sistema de classificação seguido pelo símbolo da classificação) G06T, G06K, G06V Documentação adicional pesquisada, além da mínima, na medida em que tais documentos estão incluídos nos domínios pesquisados Base de Patentes INPI-BR Base de dados eletrônica consultada durante a pesquisa internacional (nome da base de dados e, se necessário, termos usados na pesquisa) DERWENT, GOOGLE PATENTS, GOOGLE SCHOLAR		
C. DOCUMENTOS CONSIDERADOS RELEVANTES		
Categoria*	Documentos citados, com indicação de partes relevantes, se apropriado	Relevante para as reivindicações Nº
X A	AHMAD, S., and FULLER, B. " <i>Unconstrained iris segmentation using convolutional neural networks.</i> " In Asian Conference on Computer Vision, pp. 450-466. Springer, Cham, 2018. Obtido em: https://arxiv.org/pdf/1812.08245.pdf Acessado em 22/02/2022 (abstract; Seções 1, 2, 4.1)	1, 3 2, 4, 5
X A	RONNEBERGER, O., FISCHER, P., BROX, T., " <i>U-Net: Convolutional Networks for Biomedical Image Segmentation.</i> " In: NAVAB, N., HORNEGGER, J., WELLS, W., FRANGI, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham. Obtido em: https://doi.org/10.1007/978-3-319-24574-4_28 Acessado em 22/02/2022 (abstract; Seções 1, 3)	1, 2 3 – 5
☒ Documentos adicionais estão listados na continuação do quadro C ☒ Ver o anexo de famílias das patentes		
* Categorias especiais dos documentos citados: "A" documento que define o estado geral da técnica, mas não é considerado de particular relevância. "E" pedido ou patente anterior, mas publicada após ou na data do depósito internacional "L" documento que pode lançar dúvida na(s) reivindicação(ões) de prioridade ou na qual é citado para determinar a data de outra citação ou por outra razão especial "O" documento referente a uma divulgação oral, uso, exibição ou por outros meios. "P" documento publicado antes do depósito internacional, porém posterior a data de prioridade reivindicada. "T" documento publicado depois da data de depósito internacional, ou de prioridade e que não conflita como depósito, porém citado para entender o princípio ou teoria na qual se baseia a invenção. "X" documento de particular relevância; a invenção reivindicada não pode ser considerada nova e não pode ser considerada envolver uma atividade inventiva quando o documento é considerado isoladamente. "Y" documento de particular relevância; a invenção reivindicada não pode ser considerada envolver atividade inventiva quando o documento é combinado com outro documento ou mais de um, tal combinação sendo óbvia para um técnico no assunto. "&" documento membro da mesma família de patentes.		
Data da conclusão da pesquisa internacional 22 fevereiro 2022		Data do envio do relatório de pesquisa internacional: 28/02/2022
Nome e endereço postal da ISA/BR  INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL Rua Marquês de São Carlos, nº 9, 6º andar CEP: 20090-910, Centro - Rio de Janeiro/RJ Nº de fax: +55 21 3037-3663		Funcionário autorizado Fabio Pacheco Freeland Nº de telefone: +55 21 3037-3493/3742

Informação relativa a membros da família da patentes

PCT/BR2021/050478

Formulário PCT/ISA/210 (anexo da família de patentes) (Janeiro 2015)