

Visão Computacional Aplicada para Orientação de uma Cadeira de Rodas Motorizada

José Roberto de A. Aranha Junior¹, Wanderson S. da Silva¹,
Paulo N. S. do Carmo², Prof. Dr. Wener B. de Sampaio², Prof. Dr. Anselmo C. Paiva³

¹Bacharelado Interdisciplinar em Ciência e Tecnologia

²Engenharia da Computação

³Núcleo de Computação Aplicada

Universidade Federal do Maranhão (UFMA) – São Luís, MA – Brazil

{robertoaranhajr, wankapef.silva}@gmail.com,

paulonayron.s@hotmail.com, wener.sampaio@ufma.br, paiva@deinf.ufma.br

Abstract. *This work proposes a computer vision system for wheelchair control based on facial coordinates and head position estimation. This work comprises the design of a low-cost motorized wheelchair, which aims to promote independence and quality of life for people who have motor difficulties to manipulate a joystick. The tests resulted in an accuracy of 87.78 %, an precision of 90.00 % and a sensitivity of 87.70 %.*

Resumo. *Este trabalho propõe um sistema de visão computacional para controle de cadeira de rodas com base em coordenadas faciais e estimativa da posição da cabeça. Este trabalho compõe o projeto de uma cadeira de rodas motorizada de baixo custo, que tem como objetivo promover independência e qualidade de vida a pessoas que possuem dificuldades motoras para manipular um joystick. Os testes resultaram em uma acurácia de 87,78 %, a precisão de 90,00 % e uma sensibilidade de 87,70 %.*

1. Introdução

Diversos são os problemas que podem acometer o sistema locomotor humano, sendo a tetraplegia um dos mais debilitantes. Esta se refere à paralisia parcial ou completa do tronco e músculos respiratórios e dos quatro membros, sendo resultado de lesões da medula cervical [Cerezetti et al. 2012].

As pessoas que possuem lesão medular acima do nível medular T1, onde é classificada a tetraplegia, têm a mobilidade funcional comprometida, por conta da diminuição ou perda da função motora e/ou sensoria e/ou anatômica [Medola et al. 2009].

A mobilidade funcional (MF), por sua vez, é definida como a habilidade do indivíduo de se mover independentemente no ambiente, atingindo o destino desejado [Kleiner 2009].

Este trabalho propõe um sistema de Visão Computacional (VC), aliado a técnicas de Processamento Digital de Imagens (PDI), que compõem o projeto de uma cadeira de rodas motorizada, cujo objetivo é reestabelecer a MF, a fim de melhorar a qualidade de vida e promover maior independência às pessoas que possuem dificuldades de locomoção.

2. Trabalhos relacionados

Trabalhos vêm sendo desenvolvidos com a finalidade de promover melhor qualidade de vida e independência para o público alvo desse artigo.

Pinto (2017) desenvolveu um sistema de controle de cadeira de rodas de baixo custo, onde foram implementados dois modelos de interface de controle destinados a pessoas com tetraplegia. O primeiro utilizou sensores de pressão para acionar os motores por meio de sopro e sucção. O segundo modelo utilizou um joystick adaptado ao queixo para a execução dos comandos.

Rabhi *et al.* (2018) propôs um sistema de controle para cadeira de rodas inteligente de baixo custo, que usa uma câmera fixada na cadeira de rodas, capturando os movimentos das mãos do usuário em tempo real. Através de um algoritmo de VC e técnicas de inteligência artificial os movimentos são reconhecidos. Uma das principais características deste trabalho é a velocidade variável semelhante à de um joystick padrão.

3. Metodologia

A arquitetura do sistema aqui proposto segue as seguintes etapas apresentadas na Figura 1. Primeiro a aquisição, seguidos do processamento do vídeo para a detecção da face do usuário em tempo real e segmentação da mesma, que corresponde à ROI (*Region of Interest*).

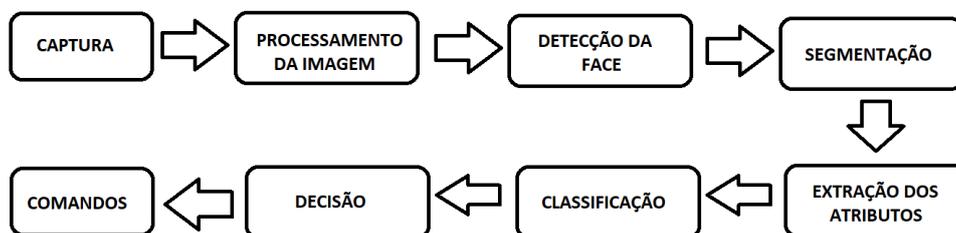


Figura 1. Arquitetura do sistema de VC.

Com a região de interesse segmentada, são encontrados pontos que servirão de atributos para as etapas de classificação, de modo a possibilitar a tomada de decisão do sistema. Nas subseções são apresentados os métodos e ferramentas utilizadas para cada etapa.

3.1. Biblioteca OpenCV

Utilizou-se a biblioteca OpenCV, referência nas áreas de VC e PDI, desenvolvida inicialmente pela Intel e atualmente de código aberto sob licença BSD [Kaehler and Bradski 2016].

O OpenCV é primordial para a aquisição e processamento da imagem, além de prover técnicas para obtenção da posição e orientação da cabeça do usuário, respectivamente descritas nas Seções 3.3 e 3.5.

3.2. Biblioteca Dlib

A Dlib contém algoritmos de aprendizado de máquinas e ferramentas de criação de softwares complexos para solucionar problemas do mundo real. É usada na indústria,

academia, além de variados domínios incluindo a robótica, dispositivos embarcados, celulares e computadores de alta performance [King 2009].

Essa ferramenta é fundamental na predição da forma e posterior detecção da face, descritas na Seção 3.4.

3.2.1. Descritores HOG - *Histogram of Oriented Gradients*

Os descritores HOG extraem características da imagem baseando-se na ideia de que objetos podem ser caracterizados de maneira satisfatória por meio da distribuição dos gradientes de intensidades dos *pixels* ou pelas direções das bordas, mesmo sem conhecer precisamente as posições de inclinação ou bordas correspondentes [Dalal and Triggs 2005].

O processo para gerar os descritores pode ser dividido em quatro etapas: cálculo do gradiente em cada *pixel*, agrupamento dos *pixels* em células, agrupamento das células em blocos e obtenção do descritor [Cruz et al. 2012].

3.3. Aquisição e Processamento de Imagem

Para a tarefa de captura da imagem utilizou-se o OpenCV. Os frames capturados por uma webcam foram redimensionados para 360p, e em seguida convertidos para a escala de cinza, de modo a reduzir o esforço computacional para as demais etapas de processamento.

3.4. Detecção de Face

Com base nos frames de entrada, ou seja, nas imagens em cinza obtidas na etapa anterior, utilizou-se um preditor de formas (*shape predictor*), cujo papel é localizar marcos faciais, que são pontos-chave de interesse ao longo da face.

A primeira etapa do processo foi a delimitação da face, isto é, a obtenção das coordenadas (x, y) do rosto na imagem. Utilizou-se para isso o algoritmo de detecção facial padrão da biblioteca Dlib. Esse algoritmo utiliza um detector de objetos, se valendo dos descritores HOG combinados ao classificador MVS Linear, descrito por Osuna *et al.* (1997), pré-treinado para rostos.

Com a face delimitada, detectou-se estruturas importantes na região. O objetivo do preditor é localizar zonas específicas da face, que são: boca, sobrancelha esquerda, sobrancelha direita, olhos, nariz e mandíbula. O preditor faz esse processo através de 68 pontos, marcados em coordenadas específicas, de regiões que envolvem cada estrutura facial. Com base nisso é feita a análise da distância provável entre os frames de entrada e o preditor, e assim um conjunto de árvores de regressão é responsável por estimar as posições de referência da face. Todas as etapas supracitadas são ilustradas na Figura 2.

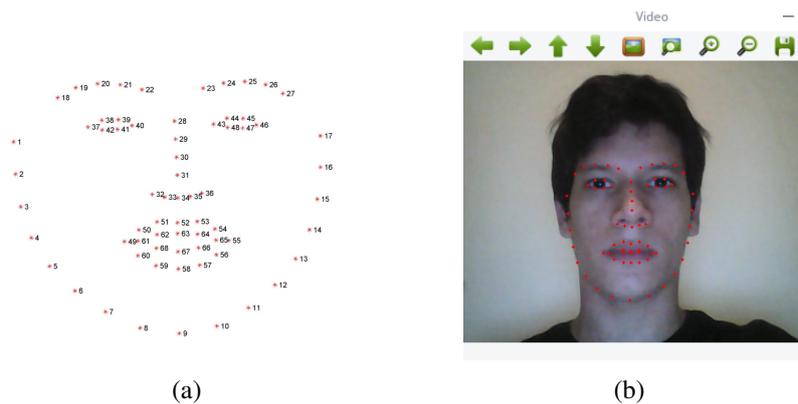


Figura 2. a) Preditor. b) Detector.

3.5. Posição estimada da cabeça

Para obter a posição e orientação da cabeça do usuário, utilizou-se a técnica de estimação de posição de objetos POSIT, proposta por DeMenthon e Davis (1995). O algoritmo relaciona os pontos 2D (Seção 3.4) e um modelo geométrico 3D aproximado da face humana. Após este relacionamento, projeta-se um vetor que representa a orientação da face, onde o mesmo é utilizado para o processo de classificação e decisão de comandos.

3.6. Classificação e Decisão de Comandos

Inicialmente, projetou-se o vetor de orientação obtido na seção 3.5, de modo que fosse possível classificá-lo através apenas de suas componentes bidimensionais.

Com este novo vetor 2D foi possível obter um retângulo cujo centro coincidissem com a origem do vetor de orientação 2D. Horizontalmente, o retângulo é equivalente ao tamanho da face detectada, e verticalmente correspondente à cerca de 40% da região delimitada, de modo a deixá-lo invariante à escala, ou seja, sem importar a distância entre o rosto e a câmera. Essa redução vertical mostrou-se necessária, uma vez que demanda-se maior liberdade de movimento, por conta das limitações do usuário.

De acordo com a posição deste retângulo menor, alinhadas lateralmente a ele, traçou-se duas retas verticais de modo a segmentar cinco regiões, denominadas comandos principais, correspondentes às seguintes direções: parado, esquerda, direita, trás e frente, conforme ilustrado na Figura 3.

Além disso, foram construídas quatro retas horizontais, também ilustradas, de modo a subdividir as regiões esquerda e direita em quatro outras subregiões, referentes aos comandos combinados, descritos como: esquerda/trás, direita/trás, esquerda/frente e direita/frente.

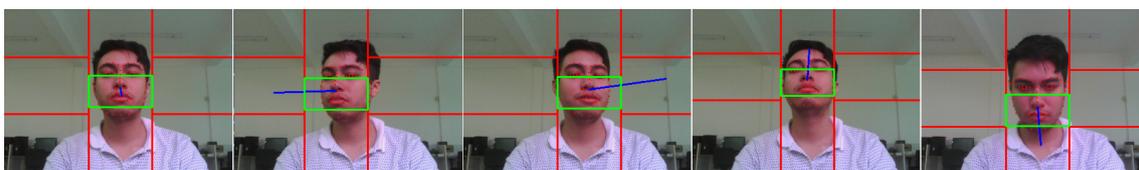


Figura 3. Regiões de Comandos.

Com base na localização do vetor nas regiões anteriormente descritas, um caractere correspondente à cada direção foi transmitido através de uma porta serial ao dispositivo embarcado utilizado na etapa de testes.

4. Etapa de Teste

Para que esta etapa fosse o mais próximo da aplicação real objetivada, que trata-se de uma cadeira de rodas, utilizou-se um sistema embarcado composto por um arduino e dois servomotores.

Foram realizados testes de interpretação dos comandos do sistema desenvolvido, de modo a testar se a captura e a resposta nos motores do embarcado condizem com a real intenção do usuário. Cada instrução foi testada 40 vezes, de modo que a transição entre os estados (direções) fosse o mais aleatória possível.

5. Resultados

Na Figura 4 é apresentada a matriz de confusão resultante dos testes realizados. Para extrair os resultados com base na matriz, utilizou-se as métricas descritas em Andrade et al. (1997). Como resultado da etapa de testes obteve-se a Acurácia igual à 87,78%, a Precisão de 90,00%, a Sensibilidade em 87,70% e uma Medida-F de 88,90%. O índice estatístico Kappa indicou que 86,25% dos comandos dados, foram interpretados corretamente.

	a	b	c	d	e	f	g	h	i	
38	1	0	0	0	0	1	0	0	0	a = Esquerda
0	37	0	0	0	0	2	0	0	1	b = Direita
0	0	22	0	0	0	1	0	0	17	c = Frente
2	0	0	35	0	0	0	0	0	3	d = Trás
5	0	1	0	32	2	0	0	0	0	e = Esquerda/Frente
1	0	1	0	0	36	0	0	2	2	f = Direita/Frente
0	0	0	2	0	0	37	1	0	0	g = Esquerda/Trás
0	0	0	1	0	0	0	39	0	0	h = Direita/Trás
0	0	0	0	0	0	0	0	0	40	i = Parado

Figura 4. Matriz resultante da etapa de testes.

Analisando a matriz de confusão (Figura 4) gerada, observa-se que os comandos de “Frente” foram os que tiveram menos resultados satisfatórios. No entanto, dos 18 casos de erro, 17 resultaram em movimento de parada e 1 de “Direita/Frente”. Ainda percebe-se que boa parcela das incursões incorretamente interpretadas, recaíram em alguma orientação semelhante ao comando de parada.

6. Conclusões e Trabalhos Futuros

Os resultados obtidos na fase de teste demonstraram que o sistema possui um bom reconhecimento dos comandos, visto que as maiores falhas na interpretação resultam em uma

orientação muito próxima à objetivada; evidenciando porém, a necessidade de melhoria da implementação do interpretador de comandos.

Para trabalhos futuros propõe-se a implementação do controle dotado de velocidade variável, para que seja semelhante à de um joystick padrão utilizado em cadeiras de rodas motorizadas, de modo que estabeleça mais conforto e segurança ao usuário. Além disso, para aumentar a acurácia e o conforto na orientação, propõe-se a utilização do filtro de Kalman, descrito por Gonzalez e Woods (2007).

Também objetiva-se a realização de testes em uma cadeira de rodas real adaptada e testes com usuários reais, com a finalidade de validar a usabilidade do sistema.

Referências

- Andrade, A., Zicker, F., and Andrade, A. (1997). Avaliação de testes diagnósticos. *Métodos de investigação epidemiológica em doenças transmissíveis*, 1:9–30.
- Cerezetti, C. R. N., Nunes, G. R., Cordeiro, D., and Tedesco, S. (2012). Lesão medular traumática e estratégias de enfrentamento: revisão crítica. *O Mundo da Saúde*, 36(2):318–326.
- Cruz, J. E., Guimarães, L. N., and Shiguemori, E. H. (2012). Um estudo da detecção automática de campos de futebol de imagens aéreas e orbitais utilizando svm e descritores hog. In *XII Workshop de Computação Aplicada. São José dos Campos:[sn]*.
- Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE.
- Dementhon, D. F. and Davis, L. S. (1995). Model-based object pose in 25 lines of code. *International journal of computer vision*, 15(1-2):123–141.
- Gonzalez, R. C. and Woods, R. E. (2007). *Digital Image Processing. Ed III*. Pearson Education Asia, New Delhi.
- Kaehler, A. and Bradski, G. (2016). *Learning OpenCV 3: computer vision in C++ with the OpenCV library*. “O’Reilly Media, Inc.”.
- King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10(Jul):1755–1758.
- Kleiner, A. F. R. (2009). Mobilidade funcional em indivíduos com paralisia cerebral espástica.
- Medola, F. O., Castello, G. L., Freitas, L. N., and Busto, R. M. (2009). Avaliação do alcance funcional de indivíduos com lesão medular espinhal usuários de cadeira de rodas. *Revista movimenta*, 2(1).
- Osuna, E., Freund, R., and Girosit, F. (1997). Training support vector machines: an application to face detection. In *Computer vision and pattern recognition, 1997. Proceedings., 1997 IEEE computer society conference on*, pages 130–136. IEEE.
- Pinto, F. F. (2017). Interfaces de controle de cadeiras de rodas motorizadas para pessoas com tetraplegia. Trabalho de conclusão de curso (graduação), Universidade de Brasília.
- Rabhi, Y., Mrabet, M., and Fnaiech, F. (2018). Intelligent control wheelchair using a new visual joystick. *Journal of Healthcare Engineering*, 2018.