TELKOMNIKA, Vol.16, No.4, August 2018, pp. 1616~1624
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v16i4.6595

Hybrid Head Tracking for Wheelchair Control Using Haar Cascade Classifsier and KCF Tracker

Fitri Utaminingrum, Yuita Arum Sari, Putra Pandu Adikara, Dahnial Syauqy, Sigit Adinugroho
Computer Vision of Research Group, Faculty of Computer Science, Universitas Brawijaya,
8 Veteran Road Malang, -65145, Indonesia
*Corresponding author, e-mail: f3_ningrum@ub.ac.id

Abstract

Disability may limit someone to move freely, especially when the severity of the disability is high. In order to help disabled people control their wheelchair, head movement-based control is preferred due to its reliability. This paper proposed a head direction detector framework which can be applied to wheelchair control. First, face and nose were detected from a video frame using Haar cascade classfier. Then, the detected bounding boxes were used to initialize Kernelized Correlation Filters tracker. Direction of a head was determined by relative position of the nose to the face, extracted from tracker's bounding boxes. Results show that the method effectively detect head direction indicated by 82% accuracy and very low detection or tracking failure.

Keywords*: Head, Detecting, Tracking*

Copyright © 2018 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

People with disability often face limitation in doing their daily activities. According to the survey [1] conducted byThe National Team for the Acceleration of Poverty Reduction (TNP2K) in 2012, 4.29% of national population suffer from various level of disability. Almost half of them have medium to severe disability level. A common way for a disable people to gain their capability in movement is to involve a wheelchair, either mechanical or motorized one. Unfortunately, not all of people cannot benefit the power of wheelchair. A people with severe disability like quadriplegia, who lost capability of controlling four limbs, cannot even move a joystick which commonly controls an electric wheelchair.

Various methods have been proposed to control a powered wheelchair. Bio generated signal such as electroencephalography [2], electromyography(EMG) [3], and electrooculography [4]. Those body-generated signals have some flaws since the acquisition of those signals is interfered by external or internal noises and devices[5]. It is also possible to steer a wheelchair using voice command [6]-[8]. Voice command may not usable for a certain category of disabilities for example mutism or people with speech disorder. It is also possible to command a wheelchair using mouth. "Sip and puff" wheelchair allows user to inhale or exhale through a straw or tube in front of a mouth to control a wheelchair [9]. There is also variation which uses wearable device placed inside a mouth, and a user is required to touch a predefined location using a tongue to signal a movement[10]. Unfortunately, any attempt to employ mouth for controlling also means disable human ability to speak and communicate effectively.

Current methods for directing an electric movement still have flaws, especially for those who have high degree of disability. Head movement is an alternative trigger for wheelchair control[11],[12],[9]. A head has more reliability to deliver control signal since it is less affected by spinal cord injury[13]. Even a large number of stroke patients who cannot move all four limbs but many of them can move their heads. Apart from reliability factor, a good control system should be non-invasive and economic [13]. In order to increase patient's comfort, computer vision based head movement detection is a feasible solution[9],[14],[15].Computer vision solution uses a digital camera to capture movement, thus the sensor is not attached to human body. The use of commercial off-the-shelf webcam or digital camera as a main sensor for head movement is a key to reduce development cost of the system.

Received June 8, 2017; Revised October 19, 2017; Accepted June 2, 2018

Head movement detection can be achieved by wide range of methods. One of them is tracking method. This method works by following the displacement of a head during its motion. This approach shows high level of accuracy [16],[17]. Features which can be used for tracking head pose are nose,mouth, and face position[18],[14]. There is also approach to use texture as position descriptor [19],[20]. However, tracking requires mandatory initialization of head position. The initialization process must also be repeated in case the tracker is lost. Haar feature-based cascade classifier provides such a function to detect a certain part of an image, including a face segment [21]. Therefore, this paper proposes a new framework to determine movement command based on head movement. Haar cascade classifiers provides initialization step to locate initial position of a head and nose. Then, a KCF based tracker continuously tracks head and nose position. Haar cascade also reinitializes overall process if tracker is lost.

2. Research Method

2.1. General Approach for Detecting Head Movement

Our head movement detection framework works by combining an objects detection followed by an object tracker. Direction is determined based on relative location of the nostril to the face. In order to prevent tracker failure, object detection is repeated after certain number of frames. Global pseudocode of our framework is summarized in Algorithm 1 and Figure 1.

Web camera is used to capture upper body person, especially upper shoulder to get the entire image of human’s head. The component for several images in computer vision is video. Therefore, getting image in real time video consist of several frames as well. After getting frame the Haar Cascade is applied, in which using two main parts of detection. There are nostril and face detection is utilized as coordinate centre to get the main feature of head movement tracking. Detection analysis is computed in order to find out the center of nose as necessary point of navigation. Detected centre of nose point is selected as the first stage before tracking. Tracking using KCF is embedded to recognize nose movement, in which head movement is depending on nostril movement.

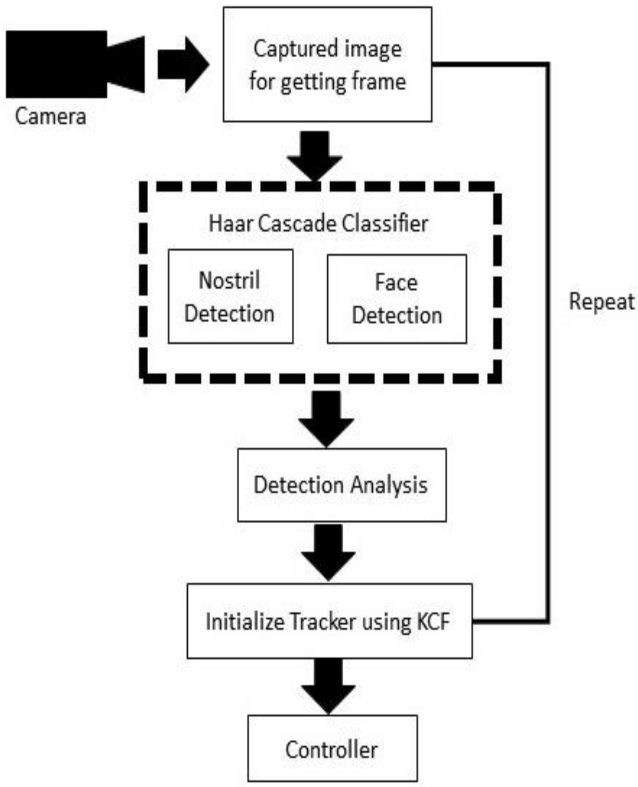


Figure 1. General method of Detecting Head Movement

ALGORITHM 1. General Pseudocode for Head Movement Detection

```
Input: frame from webcam or video file
while frame still exists:
    if frame is the first frame:
        detect nostril and face using Haar cascade classifier
            if both nostril and face are successfully detected:
                Initialize face and nostril tracker
            else:
                Redetect nostril and face
    else:
        Update face and nostril tracker
        If frame number mod 20 = 0 :
            Repeat nostril and face detection
            Initialize tracker
End
```

2.2. Detecting Frontal Face and Nostril

Frontal face and nostril acquired in order to initialized a face tracking method. Haar feature cascade classifier [22] is incorporated for that purpose. The method uses three types of kernel to extract features from an image as pictured in Figure 2. A feature is a value of subtraction result of sum of pixels under white region and sum of pixels under black one. The kernel is applied to every possible location on the image. To simplify calculation process, an integral image is built based on original image.Intensity value of integral image at position (x,y) is the sum of intensity of pixels on the left and top of (x,y) in the original image inclusive, as explained in equation 1.

$$f(x,y) = \sum_{x' \leq x, y' \leq y} f(x', y') \tag{1}$$

where $f(x,y)$ is intensity value in integral image and $f(x',y')$ is intensity value in original image

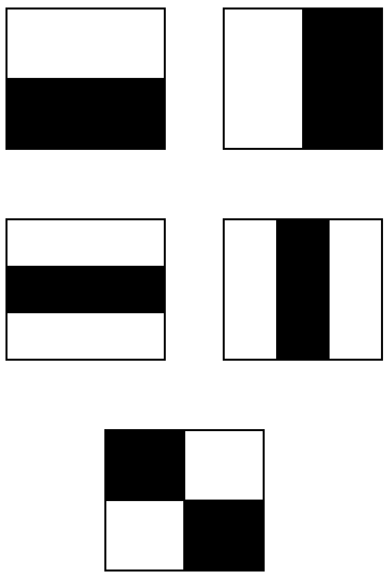


Figure 2. Haar kernels

The number of extracted features is massive and most of them are irrelevant. In order to omit non-relevant features, a form of Adaboost as shown in Equation 2 is employed. Adaboost algorithm is applied to each feature on all training images.Only features that minimize error rate are selected for testing.

$$h_j(x) = \begin{cases} 1 \ if \ p_j f_j(x) < p_j \theta_j \\ 0 \ otherwise \end{cases} \tag{2}$$

where f_j is a feature, θ_j is a threshold, and p_j is polarity showing direction of inequality sign. In testing phase, a concept of cascade classifier is applied. All features extracted during training phase are grouped into different stages and each stage consist of weak learners.Detection begins by applying each stage to label a window into positive or negative. Positive label means the window contains detected object, while negative label does not contain the object. In case of negative label, the window is slided to the next one. However, if positive label appears, classification is moved to the next stage. A window which gets all positive labels in all stages is considered as object region.

Our approach to detect nostril and frontal face uses pre-trained classifiers from Santana et al. [23] for nostril classifier and Lienhart & Maydt for frontal face classifier [24]. Face and nostril detection is implemented on Python with OpenCV library. Face and nostril detection can be used for analyzing head and eye movement systems [25].

2.3. Tracking Face and Nostril

Face and nose tracking in this paper is used Kernelized Correlation Filter (KCF) method [22], using a Dual Correlation Filter (DCF). Linear kernel is implemented in DFC. DFC with linear kernel is express ed by:

$$k(x,x') = x^T x' \tag{3}$$

and,

$$k^{xx'} = \mathcal{F}^{-1}(\textstyle\sum_c \hat{x}_c^* \odot \hat{x}_c') \tag{4}$$

denoted, $k^{xx'}$ is DFC, x is n x 1 vector that is representing a patch with the object of interest ($x=[x_1,x_2,\ldots,x_c]$) and $\mathcal{F}$ is a constant matrix and $\mathcal{F}^{-1}$ is inverse DFT.

2.4. Determining Head Direction

Direction of face at any time is estimated by examining the relative position of nose to the face. To be more precise, the center of the nose and face are compared to obtain the orientation of the head. The center of both head and nose can be easily calculated from bounding boxes as result of tracking step. Movement of a head is detected if nose's center is drifted by the center of a face by more than 1/9 of face's width. Equation 5 shows the classification of head direction based on nose and face location.

$$direction = \begin{cases} left, if\ nose's\ center > face's\ center + \frac{1}{9}\ face's\ width \\ right, if\ nose's\ center < face's\ center - \frac{1}{9}\ face's\ width \\ center, other \end{cases} \tag{5}$$

3. Results and Analysis

At the first stage of our experiment, we wanted to know if the method for detecting face works properly. To do that, we compare face detection algorithm using Haar classifier and skin classifier. The skin classifier works by determining face based on HSV thresholding. This approach is taken since the only exposed skin on our dataset is in face area.

In order to measure the performance of our method, the Boston University face database [17] is utilized. We select a fraction of videos in Boston dataset which are dominated byyaw movement. The videos used in this evaluation are jam5, jam7, jim1, jim2, jim9, and ssm8. We also limit our experiment in uniform light condition. In order to evaluate how well the methods in detecting face area, we measure the ratio of detected skin to overall pixels. Special for Haar classifier, we only detect skin area inside detected face. Figure 3 shows box plots of the ratio for jam5, jam7, jim1, jim2, jim9, and ssm8, respectively. All plots suggest that Haar classifier is more selective in selecting face area, denoted by lower value in median, lowest value, and highest value shown in the box plots. Therefore, Haar classifier is more appropriate than skin detection for selecting face area.

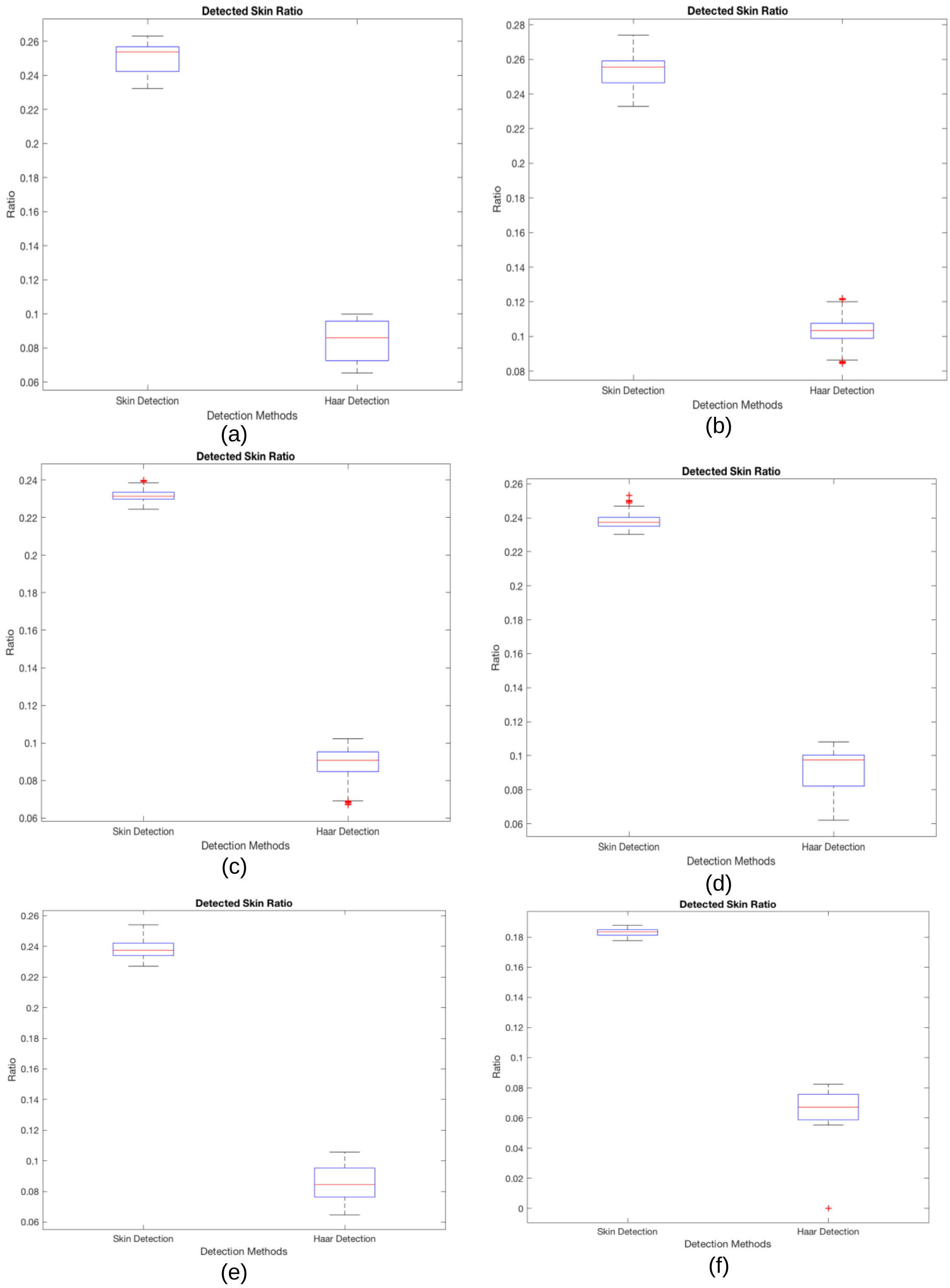
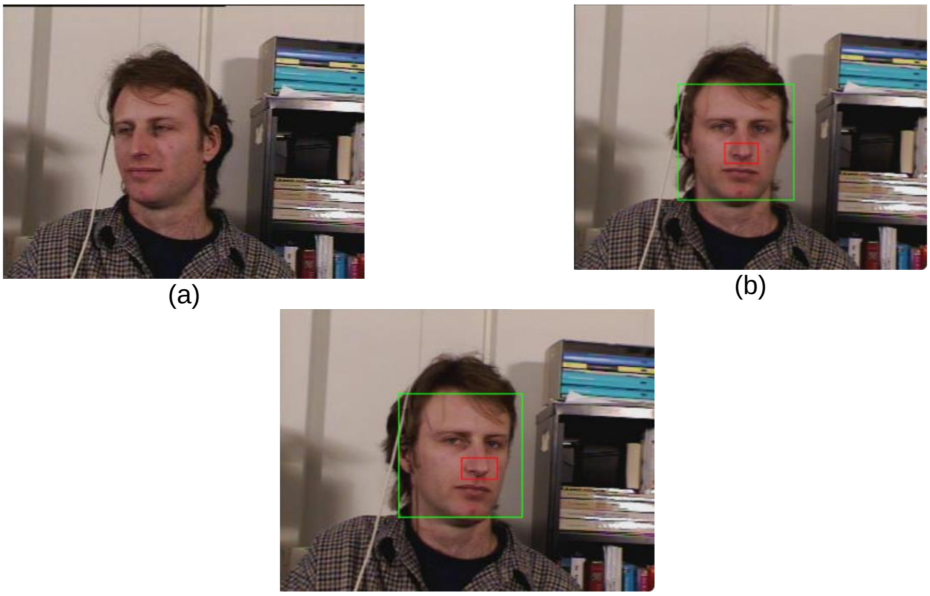


Figure 3. Box plots showing the ratio of detected face pixels to overall pixels

Further analysis reveals that skin detection is not fully appropriate for detecting face area. Figure 4 depicts detection error represented by white pixels. The error appears since the specified HSV threshold is not exclusive to skin color. In some condition, the background of the image or other objects are similar to the skin color.

Figure 4. Face detection error

The goal of this research is to provide control module for a powered wheelchair. However, in this stage we examine how good our system in detecting head direction. The dataset contains over 70 videos, each contains 200 frames, showing people with various head movement along with ground truth acquired from a Flock of Birds sensor. The ground truth data records depth, roll, yaw, and pitch. We focus our test in yaw assessment. Since the sensor acquires yaw in euler angle, we need to classify it into direction. Any movement with yaw greater than 5 degrees is classified as facing left, while movement with yaw angle less than -5 degree is considered facing right. Yaw degree between -5 and 5 degreesis labelled as facing front. A sample frame from jam7 data, detected nose and face using Haar classifier, tracking of previously detected objects shows an example of a detected yaw movement from the dataset as shown in Figure 5.



(a) (b) (c)

Figure 5. (a) A sample frame from jam7 data, (b) Detected nose and face using Haar classifier, (c) Tracking of previously detected objects
(Source: Cascia et al)

Table 1-6 show confusion matrices for jam5, jam7, jim1, jim2, jim9, and ssm8 respectively. The tables present total number of frames correctly or wrongly classified based on comparison between the output of the algorithm and ground truth. It is shown that the direction of head in most frames are correctly classified, either as center, left, or right. In case of inappropriate agreement between the result and ground truth, most cases are caused by misclassification of left and right movement as center position. There are only few evidence of misclassification of center as left or right and there is no sign of wrong classification of left as right, vice versa.

Table 1. Confusion matrix for jam5 data

		Result		
		Center	Left	Right
Ground Truth	Center	38	0	6
	Left	24	49	0
	Right	8	0	73

Table 2. Confusion matrix for jam7 data

		Result		
		Center	Left	Right
Ground Truth	Center	49	0	0
	Left	19	48	0
	Right	10	0	73

Table 3. Confusion matrix for jim1 data

		Result		
		Center	Left	Right
Ground Truth	Center	62	0	0
	Left	12	54	0
	Right	31	0	39

Table 4. Confusion matrix for jim2 data

		Result		
		Center	Left	Right
Ground Truth	Center	55	0	0
	Left	15	48	0
	Right	38	0	42

Table 5. Confusion matrix for jim9 data

		Result		
		Center	Left	Right
Ground Truth	Center	40	1	0
	Left	16	85	0
	Right	11	0	46

Table 6. Confusion matrix for ssm8 data

		Result		
		Center	Left	Right
Ground Truth	Center	29	3	0
	Left	9	86	0
	Right	11	0	61

Accuracy, defined as $accuracy = \frac{n(correctly\ classified\ frames)}{n(frames)}$, is reported in Table 7. The table reports that almost all data have accuracy of more than 80% with average accuracy of 82%. There is exception for video files jim1 and jim2, where the resultsare below 80%. We suspect that those videos only have low degree of average yaw changes that affects the selection of threshold in Equation 3. Visual inspection on those file as well as statistics from Flock of Birds sensor as served in Table 8 prove that subject on those videos only move his head slightly.

Table 7. Accuracy for each test data

	Data					
	jam5	jam7	jim1	jim2	jim9	ssm8
Accuracy	0.808	0.854	0.783	0.732	0.859	0.884

Table 8. Average yaw for each test data

	Data					
	jam5	jam7	jim1	jim2	jim9	ssm8
Average yaw	1.108	-2.034	1.147	1.173	4.305	5.950

We also conduct a test to inspect detection or tracking failure. The purpose of this test is to check if the proposed method fails to detect or track nose or face in a video frame. Table 9 presents the number of frame where detection or tracking failure occurs. The failure only happens on jam5 and jim1 videos. Further investigation reveals that the failure appears on detection phase only. There is no indication of tracking failure.

Table 9. The number of detection or tracking failure in each test data

	Data					
Number of detection	jam5	jam7	jim1	jim2	jim9	ssm8
or tracking failure	1	0	0	1	0	0

4. Conclusion

In this paper, a framework to determine head direction is introduced. The framework can be applied to a powered wheelchair as a control module. A combination of Haar cascade classifer to locate nose and head followed by KCF tracker to track both objects are the basic building of the framework. Extensive experiment on video files proves that the framework is able to recognize head direction with high accuracy as well as low detection and tracking failure. More work should be done on how to classify more precisely a nose position relative to face into a direction.

References

[1] Adioetomo SM, Mont D, Irwanto. Persons with disabilities in Indonesia : empirical facts and implications for social protection policies. Jakarta: Tim Nasional Percepatan Penanggulangan Kemiskinan. 2014.

[2] Tanaka K, Matsunaga K, Wang HO. Electroencephalogram-Based Control of an Electric Wheelchair. *IEEE Trans Robot*. 2005; 21:762-6. doi:10.1109/TRO.2004.842350.

[3] Jang G, Kim J, Lee S, Choi Y. EMG-Based Continuous Control Scheme With Simple Classifier for Electric-Powered Wheelchair. *IEEE Trans Ind Electron*. 2016; 63: 3695-705. doi:10.1109/TIE.2016.2522385.

[4] Champaty B, Jose J, Pal K, Thirugnanam A. *Development of EOG based human machine interface control system for motorized wheelchair*. 2014 Annu. Int. Conf. Emerg. Res. Areas Magn. Mach. Drives AICERAiCMMD, 2014: p. 1-7. doi:10.1109/AICERA.2014.6908256.

[5] Ruzaij MF, Neubert S, Stoll N, Thurow K. *Multi-sensor robotic-wheelchair controller for handicap and quadriplegia patients using embedded technologies*. 2016 9th Int. Conf. Hum. Syst. Interact. HSI, 2016: p. 103-9. doi:10.1109/HSI.2016.7529616.

[6] Ruzaij MF, Neubert S, Stoll N, Thurow K. *Design and testing of low cost three-modes of operation voice controller for wheelchairs and rehabilitation robotics*. IEEE 9th Int. Symp. Intell. Signal Process. WISP, 2015: p1-6. doi:10.1109/WISP.2015.7139172.

[7] Kumaran MB, Renold AP. *Implementation of voice based wheelchair for differently abled*. 2013 Fourth Int. Conf. Comput. Commun. Netw. Technol. ICCCNT, 2013: p1-6. doi:10.1109/ICCCNT.2013.6726647.

[8] Škraba A, Stojanović R, Zupan A, Kološvari A, Kofjač D. Speech-controlled cloud-based wheelchair platform for disabled persons. Microprocess Microsyst 2015; 39: 819-28. doi:10.1016/j.micpro.2015.10.004.

[9] Kupetz DJ, Wentzell SA, BuSha BF. *Head motion controlled power wheelchair*. Proc. 2010 IEEE 36th Annu. Northeast Bioeng. Conf. NEBEC, 2010: p1-2. doi:10.1109/NEBC.2010.5458224.

[10] Kim J, Park H, Bruce J, Sutton E, Rowles D, Pucci D, et al. The Tongue Enables Computer and Wheelchair Control for People with Spinal Cord Injury. *Sci Transl Med.* 2013; 5: 213ra166. doi:10.1126/scitranslmed.3006296.
[11] Kondori FA, Yousefi S, Liu L, Li H. *Head operated electric wheelchair*. Southwest Symp. Image Anal. Interpret. 2014: p. 53-6. doi:10.1109/SSIAI.2014.6806027.
[12] Ruzaij MF, Neubert S, Stoll N, Thurow K. *Design and implementation of low-cost intelligent wheelchair controller for quadriplegias and paralysis patient*. 2017 IEEE 15th Int. Symp. Appl. Mach. Intell. Inform. SAMI, 2017: p. 000399-404. doi:10.1109/SAMI.2017.7880342.
[13] Bureau M, Azkoitia JM, Ezmendi G, Manterola I, Zabaleta H, Perez M, et al. *Non-Invasive, Wireless and Universal Interface for the Control of Peripheral Devices by Means of Head Movements*. 2007 IEEE 10th Int. Conf. Rehabil. Robot. 2007: p. 124-31. doi:10.1109/ICORR.2007.4428417.
[14] Song Y, Luo Y, Lin J. *Detection of Movements of Head and Mouth to Provide Computer Access for Disabled*. 2011 Int. Conf. Technol. Appl. Artif. Intell. 2011: p. 223-6. doi:10.1109/TAAI.2011.46.
[15] Halawani A, ur Réhman S, Li H, Anani A. Active vision for controlling an electric wheelchair. Intell Serv Robot 2012; 5: 89-98. doi:10.1007/s11370-011-0098-3.
[16] Murphy-Chutorian E, Trivedi MM. Head Pose Estimation in Computer Vision: A Survey. *IEEE Trans Pattern Anal Mach Intell.* 2009; 31: 607-26. doi:10.1109/TPAMI.2008.106.
[17] Cascia ML, Sclaroff S, Athitsos V. Fast, reliable head tracking under varying illumination: an approach based on registration of texture-mapped 3D models. *IEEE Trans Pattern Anal Mach Intell*. 2000; 22: 322-36. doi:10.1109/34.845375.
[18] Jian-zheng L, Zheng Z. *Head movement recognition based on LK algorithm and Gentleboost*. 7th Int. Conf. Networked Comput. Adv. Inf. Manag., 2011: p. 232-6.
[19] Hajraoui A, Sabri M. Generic and Robust Method for Head Pose Estimation. *Indonesian Journal of Electrical Engineering and Computer Science*. 2016; 4: 439-46. doi:10.11591/ijeecs.v4.i2.pp439-446.
[20] Lin C, Qin X, Zhu G, Wei J, Lin C. Face Detection Algorithm Based on Multi-orientation Gabor Filters and Feature Fusion. *Indonesian Journal of Electrical Engineering and Computer Science*. 2013; 11: 5986-94.
[21] Utaminingrum F, Fauzi MA, Sari YA, Primaswara R, Adinugroho S. *Eye Movement as Navigator for Disabled Person*. Proc. 2016 Int. Conf. Commun. Inf. Syst., New York, NY, USA: ACM; 2016: p. 1-5. doi:10.1145/3023924.3023926.
[22] Viola P, Jones M. *Rapid object detection using a boosted cascade of simple features*. Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR 2001; 1: p. I-511-I-518. doi:10.1109/CVPR.2001.990517.
[23] Castrillón-santana M, Déniz-suárez O, Antón-canalís L, Lorenzo-navarro J. Face and Facial Feature Detection Evaluation Performance Evaluation of Public Domain Haar Detectors for Face and Facial Feature Detection. n.d.
[24] Lienhart R, Maydt J. *An extended set of Haar-like features for rapid object detection*. Proc. Int. Conf. Image Process. 2002; 1: p. I-900-I-903. doi:10.1109/ICIP.2002.1038171.
[25] Changyuan, Wang, Zhang Jing, Chen YuLong. Analyzing Head and Eye Movement System with CORBA. *Indonesian Journal of Electrical Engineering and Computer Science*. 2013; 11(11): 6618-6623.