Natural Interactions: An Application for Gestural Hands Recognition

Ricardo Proença, Arminda Guerra

Natural Interactions: An application for Gestural Hands Recognition

Ricardo Proença, Arminda Guerra

ricproenca@gmail.com; aglopes@ipcb.pt

Instituto Politécnico de Castelo Branco, Escola Superior de Tecnologia,
Av. Do Empresário, 6000 – Castelo Branco, Portugal

Abstract. This paper presents a system for the development of new human-machine interfaces focused on static gestures recognition of human hands. The proposal is to aid access to certain objects to the occupant of an intelligent wheelchair in order to facilitate their daily life. The proposed methodology relies on the use of simple computational processes and low-cost hardware. Its development involves a comprehensive approach to the problems of computer vision, based on the steps of the video image capture, image segmentation, feature extraction, pattern recognition and classification. The importance of this work relates to the need to build new models of interaction that allow, in a natural and intuitive way, to simplify the daily life of a disable person.

Keywords: Human-Computer Interaction, gesture recognition, computer vision, disabled person, assistive technology

1 Introduction

In face-to-face communication verbal and non-verbal modalities are used to complement each other. When we intend to transpose natural communication into communication with the machine, several difficulties occur, for example, those that concern user experience and, in particular, those that do not permit social inclusion.

We know that some people do not have the ability to handle some devices, but they possess enough ability to interact with a system of recognition of gestures [1]. It becomes evident the necessity to build models of Human–computer interaction through gestures that allow, in a natural and intuitive way, to these people, to easily interact with the machines.

The image acquisition and processing technologies evolution permit to build a system that enable computers to understand the sign language. Different types of algorithms for gesture recognition are being studied, but the challenge lies in the fact that this interaction becomes independent of devices that assist the identification of the human hand, such as gloves or sensors for motion capture.

The proposed work is the development of a system for static gesture recognition with one hand only, located in front of a webcam, in a simple and uniform background scenario, without the support of assistive materials. This interface could be applied to an intelligent wheelchair in order to facilitate disabled peoples' quality of life.

2 Related Work

Several gestural proposals for various areas, such as entertainment, medicine, marketing, smart home control, elderly and disable care [2], [3] as well as intelligent wheelchairs [4], [5], [6], [7], [8] were found in literature review. However, the majority of studies are exclusively related to the chair's movement and we did not find the complete mobility concerns. We also noticed the existence of wheelchair-mounted robotic arm applications [9], [10], [11], [12] that uses different interaction devices (such as joystick), but none so intuitive that could be considered natural (to humans).

In fact, this relation is not fully effective to personal conflict resolution without considering mobility. Some people, for example, do not hold the necessary dexterity to handle these devices, but they have enough ability to interact with a system for gesture recognition [1] that certainly minimizes many of the difficulties and inhibitions.

3 The Context: Assistive Technology

We are daily faced with several examples of people who suffer road accidents. The consequences fully affect the mobility of the individual either physically, psychologically and socially. This situation promotes numerous side effects, some of which demand alternative means of mobility such as wheelchairs.

The increasing number of people with mobility disabilities requires assistive technologies to provide their quality of life improvement. However, there are many obstacles to use technology solutions because they require expensive resources or further studies in order to improve every citizen access.

4 Proposed Solution

We considered the case of an individual who needs to have access to an object placed on a surface with a large area, inhibitory of normal and autonomous actions.

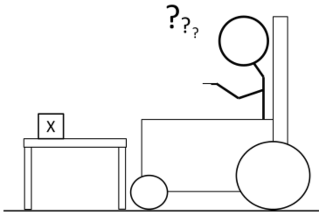

Fig. 1. Problematic situation

It is our understanding, according the research we did, that image acquisition and processing technologies' evolution, allows building a computational system that makes the machine to recognize human gestures. However, the construction of this system requires the use of various programming algorithms quite complexes. We list general procedures that entail: (1) image capture, (2) human hand segmentation, (3) feature extraction, (4) classification of gestures.

We considered the gestural algorithms development, based on low cost webcams, a great stimulus in Human-Machine Interaction.

Given the results we obtained in the course of this investigation, we propose a system to help the occupant of a wheelchair to interact with the world, especially, to have access to certain objects in order to facilitate their daily life.

In terms of hardware, the prototype consists of an intelligent wheelchair, a laptop with a built-in webcam and a robotic arm. As to software, this system consists in static gesture recognition, performed in real time, through a hand located in front of a webcam.

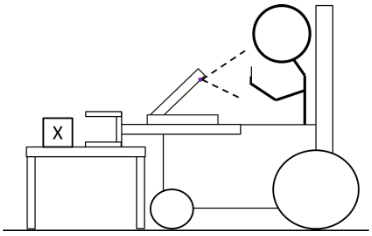

Fig. 2. Proposed system

Our concept of mobility will be based on the transposition of each executed gesture to a robotic arm action. This transposition is set from 6 gestures that we present in fig. 3, fulfilling each a specific function in the context of facilitating the mobility of the person: (a) pause; (b) stretch mechanical arm with speed 1 (c) stretch mechanical arm with speed 2 (d) open claw, (e) close claw; (f) move arm to the starting position.

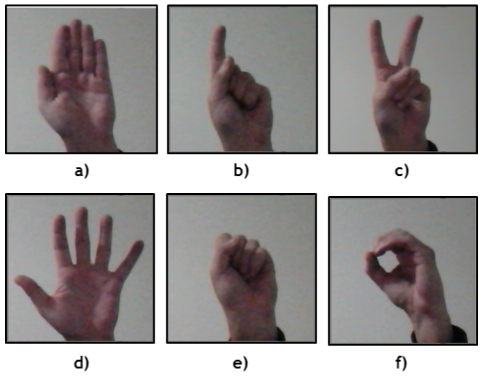

Fig. 3. Hand recognition gestures

With this method, we intend to show that it is possible to construct a human-machine interaction system, facilitating the life of the person with inhibitions, with an acceptable level of environmental restrictions.

Again, we can prove that the fact that a person has to use a glove, or any other device that aids the identification of the human hand, is inhibiting the free, systematic and autonomous individual action.

The scope of this work is limited to the recognition of gestures; we want to identify the static gesture shown by the occupant of the wheelchair, between a set of predefined gestures system. The prototype that we outlined has not been built so far.

Finally, we launched the challenge to awake attentions and possible interests for the implementation of this type of equipment that will allow, in our view, more disinhibiting action of the individual, because it determines the elements that are not always available to anyone of us, even in conditions of normal mobility.

5 Natural Interactions

Natural interaction is considered through the way people naturally communicate through gestures, expressions, movements, and see the world by looking around and manipulating physical artifacts; the key assumption is that people should be allowed to interact with technology as they are used to interact with the real world in everyday life.

Our focus is on gestural communication. The gestures can naturally increase the interaction between the user and the computer, replacing devices such as the mouse, keyboard, joystick or buttons in machines control. However, a great incitement for gestural interfaces development comes from the growth of applications for virtual environments [13], such as surgery simulation. In the majority of these applications, the gestures are seen as triggers to virtual objects control, simulating the 2D, 3D or abstractions of real objects, such as robotic arms.

5.1 Gestures

There are many definitions and debates about the meaning of gesture. Pavlovic [14] presents a definition of hand gesture that combines the development of posture and position in time, with the intention of communication or manipulation.

In Human–computer interaction field, gestures imply the use of the hands to instruct a machine. Its meaning depends on the system that integrates and evaluates them. Point or rotate actions are representative actions of this type of gestures.

We can differentiate the gestures in two types: dynamic and static. The dynamic gestures, necessarily, involve time evolution [15] and require movement to transmit the intended message. The gesture to say "goodbye" is an example of a dynamic gesture performed with an open hand and is only meaningful if done with movement.

Furthermore, static gestures do not include movement to transmit the message. The same static open hand, with no motion made, forward the information to "stop".

In Computer Vision, we can state that static gesture is a configuration of the hand pose represented as a single image. A dynamic gesture implies movement, represented by a sequence of images.

However, in this work, we will limit ourselves to the static symbolic gestures performed by one hand.

5.2 Static Gesture Recognition

Static gesture is represented by an image; the recognition is based on feature extraction discriminants of the image. We can, for instance, to extract features from the directions of the fingers or the contours of the hand.

The position of the palm of the hand and of the fingers has been widely used for gesture recognition [16], [17]. We can also treat other information about the hands as the angles [18], roughness [19] or sampling points on the boundary region of the hand [20], [21].

Another category of resources abundantly used implies the notion of moment. It is a weighted average of the pixels intensities of an image, usually chosen to have some interpretation. We can find statistical moments [22], Zernike moments [23] and Hu moments [24], which are more used.

Fourier analysis can also be applied for representing the limits of the hand [25].

The feature extraction is a complex computational problem because it requires perfect background characteristics and high computational resources. Disorganized background images with changes in lighting conditions become very difficult to handle.

Allied to this, the used algorithms require loading the whole image into memory, which complicates the computer resources management.

After extracting the features, then we can classify the gestures. The more simplistic way of static hand recognition is to count the number of fingers [17]. Nevertheless, recognition can be seen as a matching process, applying cluster algorithms [25], [26]. In this case, recognition consists of finding the best match between images that represent static gestures. We may also do the same correspondence with Chamfer distance algorithm [27]. In addition, the Neural Networks are widely used in the hand postures detection and classification [28], using two or three layers of neurons. More recently, techniques of face detection have been applied to this problem. The Viola and Jones system that leverages the features Haar [29] is a consensus for subject researchers.

6 Interface Design: Process Description

Based on [30], the proposed method for symbolic gesture recognition follows a logical and complementary tasks' sequence in terms of digital image processing. The fundamental steps for the proposed work are presented on fig.4.

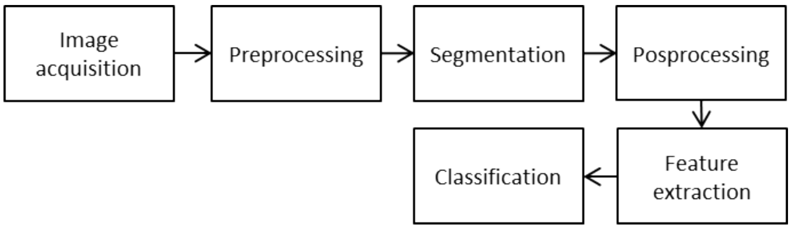

Fig. 4. Fundamental steps for the proposed work

The image acquisition is made by a low cost webcam. The preprocessing and post processing use image enhancement and noise reduction techniques by analyzing environmental lightning conditions and background colors. This happens due to difficulties in illumination. The image segmentation obtained by Computer Vision algorithms, allows the separation of full human hand in a simple and uniform background scenario, without benefit of supporting devices. The features to extract are based in the geometric shape of the hand, fast enough to be applied in real time. The classification step is based on the use of the recognition patterns classifier for training and to classify the input data.

6.1 Image Acquisition

The process starts with the gesture acquisition by the webcam and presented through video images (fig. 5).

Fig. 5. Captured frame

6.2 Segmentation

Segmentation is the main part of the entire process and it assumes a sequence of tasks (Fig. 6). This stage makes the human hand recognition by skin color detection.

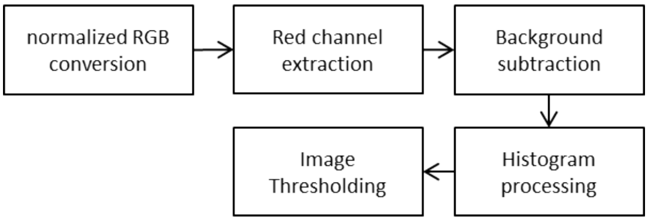

Fig. 6. Key steps of segmentation

The procedure starts with the frames conversion captured for the normalized RGB color space. Knowing that the majority of human skin color tends to cluster in the red channel, we extract this channel and then, we subtract the background of the extracted red channel image. Background subtraction allows us to obtain a greyscale image with color intensity changes between the background and the hand, as shown in (fig. 7).

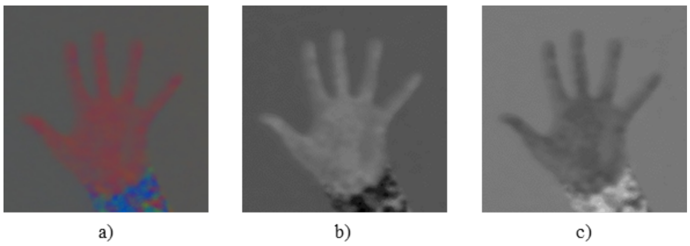

Fig. 7. – a) Normalized RGB image b) Red channel extracted c) Background subtracted

Then, we design the image histogram and compute the optimal thresholding value. We used the Method of the Valley that consists in checking how many regions there are (peaks and valleys) in the histogram and use the valley information to set the value of the threshold. Fig. 8 shows the histogram and optimal threshold value identified.

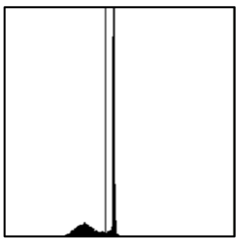

Fig. 8. Image Histogram with optimum threshold identified

Finally, we use the found threshold value to binarize the image (fig. 9).

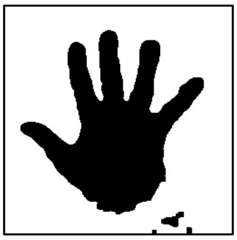

Fig. 9. Binary image

6.3 Feature Finding

The previous step provides an image where we can find the edge pixels that separate the hand from the background, but without information about the edges as entities in themselves, thus, the next step is to be able to gather the whole pixels on the edge contours surrounding the stain of segmented hand.

Fig. 10 shows that trough the extracted contour we can calculate several structures such as convex hull, contour edges, fingers and surrounding rectangles.

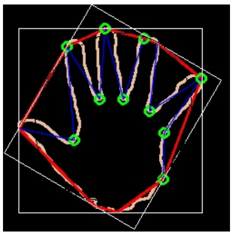

Fig. 10. Calculated structures to feature extraction

6.4 Feature extraction

The feature extraction lets obtaining a set of feature vectors, also called descriptors that can distinguish, accurately, each hand gesture. Our approach proposes the use of geometric features with Hu invariant moments [24]. The presented features are the most discriminatory from the six gestures chosen: contour size; contour circularity; contour eccentricity; convex hull size; bounding box proportion; bounding box area; rotated bounding box area; main inertia axe; Hu Moments.

Some problems during the gesture recognition may occur if there is a large rotational movement, because the shape and proportions of features that represent the gesture may change.

We noticed that these features are easy to extract, are independent and are sufficient to distinguish the amount of gestures chosen by us, however, these are not suitable for all types of gestures.

6.5 Classification

For pattern recognition and classification, we choose the Support Vector Machines (SVM) machine learning technique that has been recognized in recent years. This technique is used in various pattern recognition tasks, always with results superior to those achieved by similar techniques in many applications. Its use comes from some of its main characteristics [31]:

- Good generalizability;
- Robustness in larger dimensions;
- Well-defined theory in Mathematics and Statistics.

7 Tests

The system performs well as long as the skin-like colors in the background do not exist and if it is tested indoor. Results pointed for a good removal of luminosity influence during the segmentation process, so it becomes less dependent on the lighting conditions, which has always been a critical obstacle to the image recognition. The system operates in real time with around 30 frames per second, requiring an effort to minimize time processing of the tasks.

We saved 100 images representing each gesture to our database. For testing we have also saved 100 images for each gesture of 5 different people.

Table 1 summarizes the results of the hand classification tests.

Table 1. Gesture classification results

Gestures	Frames	Correct	Incorrect	% Correct	% Incorrect
A, B, C, D, E, F	3000	2974	26	99.1%	0.9%

8 Tasks and Activities to Build the Prototype

The main goal of the proposed work was the development of a system for recognizing static gestures of human hands of different people, situated in front of a webcam.

The choice of appropriate technology for this type of system could limit the development and visualization of the results.

This section presents and clarifies the work context for the development of the proposed system. Considering the literature review process carried out, we discovered interesting and innovative approaches for the application's development based on the gestural recognition. However, we found an absence of this type of systems for assistive technologies regarding wheelchair intelligent robotic arms.

Based on the lack of information, we create our design process for the proposed system. The decision about the software and hardware to be used on the development of our system was the main concern. Conversely, we needed to consider the basic principles defined previously: the system should operate in real time with minimal restrictions and the system should have a low cost demand in order to be applied in everyday life by a significant number of users who needed it.

Based on these decisions fig. 11 presents the first phase of the design process:

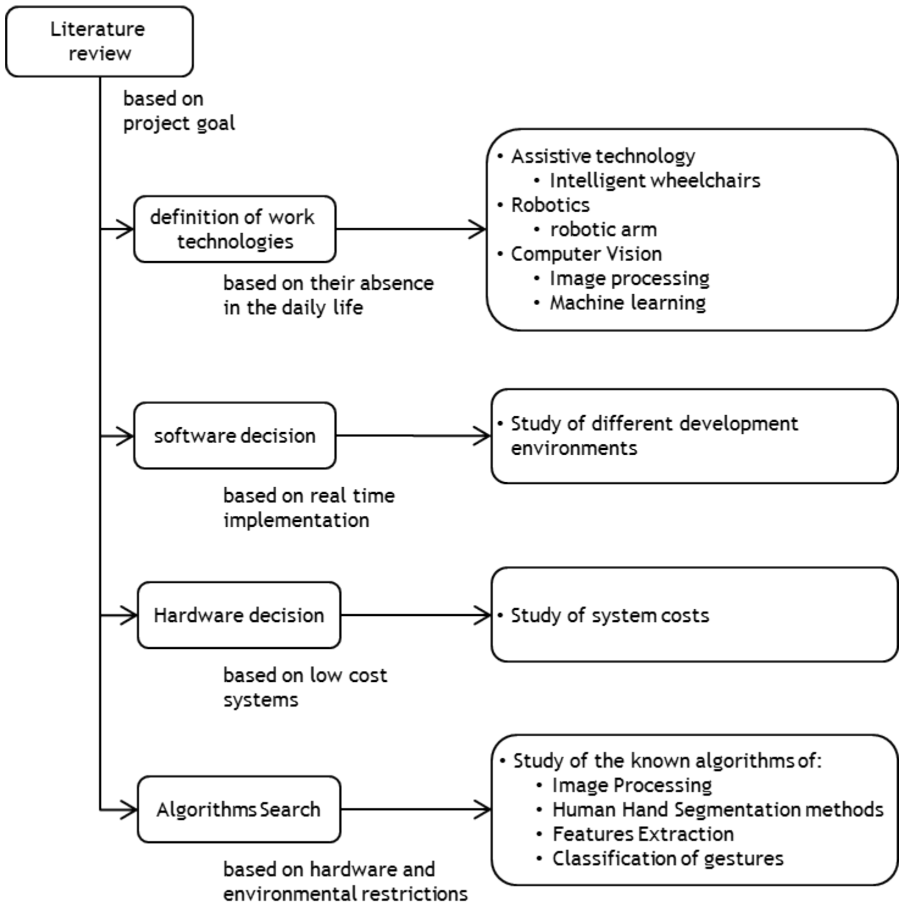

Fig. 11. The design process – first phase

After acquiring the theoretical knowledge about the subject, on literature review process, based on our project goals, the next step was the definition of the technologies that should be chosen considering the facility of their use in a daily life purpose by all kind of users. An intelligent wheelchair with a robotic arm and using image processing technology was the defined goal. Then, the main argument for the choice of the software was made after the study of different development environments based on real time implementation. The study of the system cost, the lowest one, contributed to the hardware selection to be used. Finally, we searched for the most used algorithms for each experiment steps (image processing, human hand segmentation methods, feature extraction and gestures classification) taking into consideration environment restrictions.

The relationship between quality and performance for the real-time system was taken into account.

Fig. 12 shows the implementation task process referring the decision taken during each task: image acquisition, color space conversion, histogram processing, hand segmentation, feature finding, feature extraction, classification of gestures and system tests.

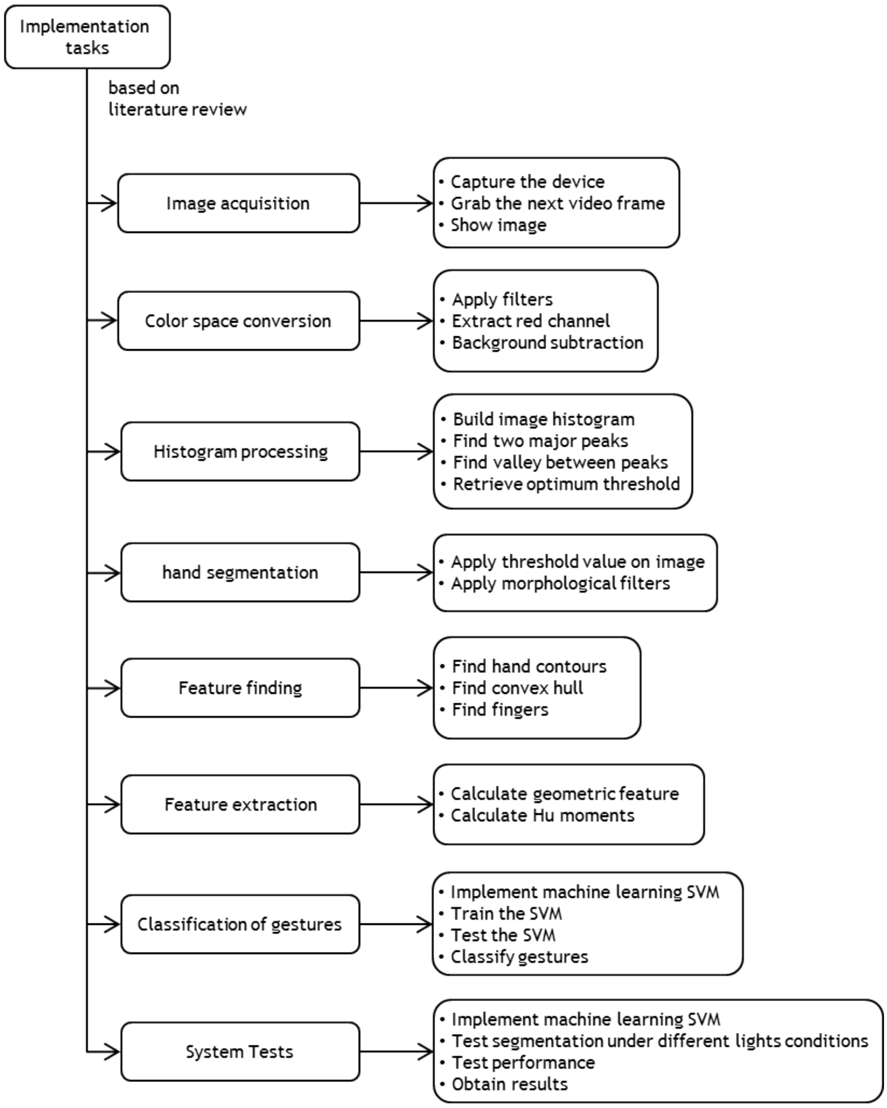

Fig. 12. Implementation task process

Some potential users of the system, around twelve, were, informally, interviewed, at the beginning of this project and later on a prototype phase. They considered that the proposed solution will be important since it preview affordance for a highest number of people who have social, financial or accessibility restrictions.

9 Conclusions

It has been shown that it is possible to interact with a machine naturally and intuitively through hand gestures without requiring support material such as gloves or markers. And that there is a possibility to develop Computer Vision interactive systems based and implemented on conventional computers using inexpensive devices, accessible to all, facilitating, especially disable mobility people with an acceptable level of environmental restrictions.

The proposed system produces satisfactory results with fixed constant lighting conditions, with a non-skin color scenario. However, the method used for skin segmentation exposes some flaws, mainly caused due to lighting variations causing glare, reflections and shadows. These shortcomings are minimized applying morphological filters and smoothing filters.

Conversely, the systems works with quite different hand gestures, but it shows some weakness when using more identical gestures such as sign language. Thus, the scope of this work is limited to the recognition of gestures; we wanted to identify the static gesture shown by the occupant of the wheelchair. The prototype that we outline has not been built so far.

In future work, the study of methods that may respond better to lighting variations, to different scenarios and how to be more effective in identifying the skin, even from other races would be helpful. The gesture recognition through dynamic movement analysis can also be a target for further investigation. However, we launched the challenge to awake attentions and possible interests for the implementation of this type of equipment that will be, in our point of view, more disinhibiting action of the individual, because it determines the element that are not always available to anyone.

10 References

[1] R. Pausch and R. D. Williams, "Giving CANDY to Children: User-Tailored Gesture Input Driving an Articulator-Based Speech Synthesizer," *Communications of the ACM,* pp. 58-66, 2012 5 5.

[2] M. Lande and M. Samvatsar, "Multimodal System using Human Computer Interface," *International Journal of Engineering and Innovative Technology (IJEIT),* vol. 1, 2012.

[3] S. Chhabria and R. Dharaskar, "Multimodal Interface for Disabled Persons," *International Journal of Computer Science and Communication V,* pp. 223-228, 2012.

[4] Y. Zhang, Z. J. and Y. Luo, "A novel intelligent wheelchair control system based on hand gesture recognition," in *IEEE/ICME International Conference on Complex Medical Engineering (CME)*, 2011.

[5] Y. Kobayashi, Y. Kinpara, T. Shibusawa and Y. Kuno, "Robotic wheelchair based on observations of people using integrated sensors," *IROS'09 Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems,* pp. 2013-2018, 2009.

[6] R. Simpson, E. LoPresti, S. Hayashi, I. Nourbakhsh and D. Miller, "The smart Wheelchair component system," *Journal of Rehabilitation Research and Development,* vol. 41, pp. 429-442, 2004.

[7] P. Jia, H. Hu, T. Lu and K. Yuan, "Head Gesture Recognition for Hands-free Control of an Intelligent Wheelchair," *Industrial Robot: An International Journal,* vol. 34, pp. 60-68, 2007.

[8] K. Kumar and M. Dinesh, "Hand Gesture Recognition Wheelchair Control by Digital Image Processing Technique," *International Journal of Communications and Engineering,* vol. 4, pp. 67-71, 2012.

[9] M. Hillman, K. Hagan, S. Hagan, J. Jepson and R. Orpwood, "The Weston Wheelchair Mounted Assistive Robot - The Design Story," *Robotica,* vol. 20, pp. 125-132, 2002.

[10] K. Tsui, H. Yanco, D. Kontak and L. Beliveau, "Development and Evaluation of a Flexible Interface for a Wheelchair Mounted Robotic Arm," in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, 2008.

[11] M. Palankar, K. De Laurentis, R. Alqasemi, E. Veras, R. Dubey, Y. Arbel and E. Donchin, "Control of a 9-DoF Wheelchair-mounted robotic arm system using a P300 Brain Computer Interface: Initial experiments," in *Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics*, 2009.

[12] Y. Touati and A. Ali-Cherif, "Smart Powered Wheelchair Platform Design and control for People with Severe Disabilities," *Software Engineering,* pp. 49-56, 2012.

[13] W. M. Krueger, Artificial Reality II, Addison-Wesley, 1991.

[14] V. I. Pavlovic, R. Sharma and T. S. Huang, "Visual Interpretation of Hand Gestures for Human Computer Interaction: A Review," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997.

[15] H. M. T. B. &. M. C. B. Birk, "Real-Time Recognition of Hand Alphabet Gestures Using Principal Component Analysis," in *10th Scandinavian Conference on Image Analysis*, 1997.

[16] J. MacLean, C. Pantofaru, L. Wood, R. Herpers, K. Derpanis, D. Topalovic and J. Tsotsos, "Fast Hand Gesture Recognition for Real-Time Teleconferencing Applications," in *IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, Vancouver, 2001.

[17] H. Kim and D. W. Fellner, "Interaction with Hand Gesture for a Back-Projection Wall," in *CGI '04 Proceedings of the Computer Graphics International*,

Washington, 2001.

[18] Y. Sato, M. Saito and H. Koik, "Real-Time Input of 3D Pose and Gestures of a User's Hand and Its Applications for HCI," in *2001 IEEE Virtual Reality Conference*, Yokohama, 2001.

[19] H. Jang, D. Jun-Hyeong , J. Jin-Woo and Z. Z. Bien, "Two-staged hand-posture recognition method for softremocon system," in *Systems, Man and Cybernetics, 2005 IEEE International Conference on, Volume: 1*, 2005.

[20] Y. Hamada, N. Shimada and Y. Shirai, "Hand Shape Estimation Using Sequence of Multi-Ocular Images Based on Transition Network," in *International Conference on Vision Interface*, 2002.

[21] Y. Hamada, N. Shimada and Y. Shirai, "Hand shape estimation using image transition network," in *HUMO '00 Proceedings of the Workshop on Human Motion*, 2000.

[22] L. K. Lee, S. Kim, Y. K. Choi and M. H. Lee, "Recognition of hand gesture to human-computer interaction," in *26th Annual Conference of the IEEE Industrial Electronics Society, vol. 3*, 2000.

[23] Y. Sribooruang, P. Kumhom and K. Chamnongthai, "Hand posture classification using wavelet moment invariant," in *Virtual Environments, Human-Computer Interfaces and Measurement Systems*, 2004.

[24] M. K. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Trans. Info. Theory,* Vols. IT-8, pp. 179-187, 1962.

[25] A. Licsár and T. Szirányi, "Dynamic training of hand gesture recognition system," in *17th International Conference on Pattern Recognition (ICPR'04) - Volume 4*, 2004.

[26] H. Zhou, D. J. Lin and T. S. Huang, "Static hand gesture recognition based on local orientation histogram feature distribution model," in *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Volume 10*, 2004.

[27] V. Athitsos and S. Sclaroff, "An appearance-based framework for 3d hand shape classification and camera viewpoint estimation," in *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 2001.

[28] S. Marcel, O. Bernier, J.-E. Viallet and D. Collobert, "Hand gesture recognition using inpu/output hidden markov models," in *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000.

[29] M. J. Jones and P. Viola, "Face recognition using boosted local features," in *International Conference on Computer Vision*, 2003.

[30] R. Gonzalez and R. Woods, Digital Image Processing, 2nd Edition, Addison Wesley, 2002.

[31] A. Smola, "Geometry and invariance in kernel based methods.," in *Advances in Kernel Methods: Support Vector Learning*, MIT Press, 1999, pp. 89-116.