

Vision-based Portuguese Sign Language Recognition System

Paulo Trigueiros^{1,2,3,4}, Fernando Ribeiro^{2,3} and Luís Paulo Reis^{3,4,5}

¹ Instituto Politécnico do Porto, Porto, Portugal

² Departamento de Electrónica Industrial, Escola de Engenharia da Universidade do Minho, Campus de Azurém 4800-058, Guimarães, Portugal

³ Centro Algoritmi, Escola de Engenharia da Universidade do Minho, Guimarães, Portugal

⁴ Laboratório de Inteligência Artificial e Ciência de Computadores, Porto, Portugal

⁵ Departamento de Sistemas de Informação, Escola de Engenharia da Universidade do Minho, Campus de Azurém 4800-058, Guimarães, Portugal

`pjt@iscap.ipp.pt, fernando@dei.uminho.pt, lpreis@dsi.uminho.pt`

Abstract. Vision-based hand gesture recognition is an area of active current research in computer vision and machine learning. Being a natural way of human interaction, it is an area where many researchers are working on, with the goal of making human computer interaction (HCI) easier and natural, without the need for any extra devices. So, the primary goal of gesture recognition research is to create systems, which can identify specific human gestures and use them, for example, to convey information. For that, vision-based hand gesture interfaces require fast and extremely robust hand detection, and gesture recognition in real time. Hand gestures are a powerful human communication modality with lots of potential applications and in this context we have sign language recognition, the communication method of deaf people. Sign languages are not standard and universal and the grammars differ from country to country. In this paper, a real-time system able to interpret the Portuguese Sign Language is presented and described. Experiments showed that the system was able to reliably recognize the vowels in real-time, with an accuracy of 99.4% with one dataset of features and an accuracy of 99.6% with a second dataset of features. Although the implemented solution was only trained to recognize the vowels, it is easily extended to recognize the rest of the alphabet, being a solid foundation for the development of any vision-based sign language recognition user interface system.

Keywords: Sign Language Recognition, Hand Gestures, Hand Postures, Gesture Classification, Computer Vision, Machine Learning.

1 Introduction

Hand gesture recognition for human computer interaction is an area of active research in computer vision and machine learning. One of its primary goals is to create systems, which can identify specific gestures and use them to convey information or

to control a device. Though, gestures need to be modelled in the spatial and temporal domains, where a hand posture is the static structure of the hand and a gesture is the dynamic movement of the hand. There are basically two types of approaches for hand gesture recognition: *vision-based approaches* and *data glove approaches*. This work main focus is on creating a vision-based system able to do real-time sign language recognition. The reason for choosing a system based on vision relates to the fact that it provides a simpler and more intuitive way of communication between a human and a computer. Being hand-pose one of the most important communication tools in human's daily life, and with the continuous advances of image and video processing techniques, research on human-machine interaction through gesture recognition led to the use of such technology in a very broad range of applications, like touch screens, video game consoles, virtual reality, medical applications, and sign language recognition. Although sign language is the most natural way of exchanging information among deaf people it has been observed that they are facing difficulties with normal people interaction. Sign language consists of vocabulary of signs in exactly the same way as spoken language consists of a vocabulary of words. Sign languages are not standard and universal and the grammars differ from country to country. The Portuguese Sign Language (PSL), for example, involves hand movements, body movements and facial expressions [1]. The purpose of Sign Language Recognition (SLR) systems is to provide an efficient and accurate way to convert sign language into text or voice has aids for the hearing impaired for example, or enabling very young children to interact with computers (recognizing sign language), among others. Since SLR implies conveying meaningful information through the use of hand gestures [2], careful feature selection and extraction are very important aspects to consider. Since visual features provide a description of the image content [3], their proper choice for image classification is vital for the future performance of the recognition system. Viewpoint invariance and user independence are two important requirements for this type of systems. In this paper, a vision-based system able to interpret static hand gestures from the Portuguese Sign Language alphabet is presented and described. The selected hand features were analysed with the help of Rapid Miner [4] in order to find the best learner for the problem under study. The rest of the paper is as follows. First we review related work in section **Erro! A origem da referência não foi encontrada.** Section **Erro! A origem da referência não foi encontrada.** presents the Sign Language Recognition prototype architecture and implementation. In section 3.2 the experimental methodology is describes and the obtained results are presented and discussed. Conclusions and future work are drawn in section **Erro! A origem da referência não foi encontrada.**

2 Related Work

Hand gesture recognition, either static or dynamic, for human computer interaction in real time systems is a challenging task and an area of active research with many possible applications. There are many studies on gesture recognition and methodologies well presented in [5, 6]. As explained in the previous section, careful hand features selection and extraction are very important aspects to consider in

computer vision applications for hand gesture recognition and classification for real-time human-computer interaction. This step is crucial to determine in the future whether given hands shape matches a given model, or which of the representative classes is the most similar. According to Wacs [7] proper feature selection, and their combination with sophisticated learning and recognition algorithms, can affect the success or failure of any existing and future work in the field of human computer interaction using hand gestures. Trigueiros, in his study [8], presented a comparative study of seven different algorithms for hand feature extraction with the goal of static hand gesture classification. The results showed that the radial signature and the centroid distance were the features that when used separately obtained better results, being at the same time simple in terms of computational complexity. He has also implemented a vision-based system [9], able to drive a wheelchair with a minimum number of finger commands. In the system, the user hand is detected and segmented, and fingertips are extracted and used as features to build user commands for wheelchair control. Wang [10] used the discrete Adaboost learning algorithm integrated with SIFT features for accomplishing in-plane rotation invariant, scale invariant and multi-view hand detection. Conceil [11] compared two different shape descriptors, Fourier descriptors and Hu moments, for the recognition of 11 hand postures in a vision based approach. They concluded that Fourier descriptors gives good recognition rates in comparison with Hu moments. Barczak [12] performed a performance comparison of Fourier descriptors and geometric moment invariants on an American Sign Language database. The results showed that both descriptors are unable to differentiate some classes in the database. Bourennane [13] presented a shape descriptor comparison for hand posture recognition from video, with the objective of finding a good compromise between accuracy of recognition and computational load for a real-time application. They run experiments on two families of contour-based Fourier descriptors and two sets of region based moments, all of them invariant to translation, rotation and scale-changes of hands. They performed systematic tests on the Triesch benchmark database [14] and on their own with more realistic conditions, as they claim. The overall result of the research showed that the common set Fourier descriptors when combined with the k-nearest neighbour classifier had the highest recognition rate, reaching 100% in the learning set and 88% in the test set. Huynh [15] presents an evaluation of the SIFT (scale invariant feature transform), Colour SIFT, and SURF (speeded up robust features) descriptors on very low resolution images. The performance of the three descriptors is compared against each other on the precision and recall measures using ground truth correct matching data. His experimental results showed that both SIFT and colour SIFT are more robust under changes of viewing angle and viewing distance but SURF is superior under changes of illumination and blurring. In terms of computation time, the SURF descriptors offer themselves as a good alternative to SIFT and CSIFT. Fang [16] to address the problem of large number of labelled samples, the usually costly time spent on training, conversion or normalization of features into a unified feature space, presented a hand posture recognition approach with what they called a co-training strategy [17]. The main idea is to train two different classifiers with each other and improve the performance of both classifiers with unlabelled samples. They claim that their method improves the recognition performance with less labelled data in a semi-supervised way. Rayi [18] used the centroid distance Fourier descriptors as hand

shape descriptors in sign language recognition. Their test results showed that the Fourier descriptors and the Manhattan distance-based classifier achieved recognition rates of 95% with small computational latency. In terms of classification and learning, machine learning algorithms have been applied successfully to many fields of research like, face recognition and facial expressions [19, 20], automatic recognition of a musical gesture by a computer [21], classification of robotic soccer formations [22], classifying human physical activity from on-body accelerometers [23], automatic road-sign detection [24, 25], static hand gesture classification [26], serious games applied to rehabilitation [27-30] and intelligent wheelchairs [20, 31-34]. Trigueiros [26] have made a comparative study of four machine learning algorithms applied to two hand features datasets. In their study the datasets had a mixture of hand features. Ke [35] used a Support Vector Machine (SVM) in the implementation of a real-time hand gesture recognition system for human robot interaction. Masaki [36] used a SVM in conjunction with a SOM (Self-Organizing Map) for the automatic learning of a gesture recognition mode. He first applies the SOM to divide the sample into phases and construct a state machine, and then he applies the SVM to learn the transition conditions between nodes. Almeida [37] proposed a classification approach to identify the team's formation in the robotic soccer domain for the two dimensional (2D) simulation league employing Data Mining classification techniques. Hidden Markov Models (HMMs) have been widely used in a successfully way in other areas of application, and also applied quite successfully to gesture recognition. Oka [38] developed a gesture recognition system based on measured finger trajectories for an augmented desk interface system. They have used a Kalman filter for the prediction of multiple finger locations and an HMM for gesture recognition. Perrin [39] described a finger tracking gesture recognition system based on a laser tracking mechanism which can be used in hand-held devices. They have used HMM for their gesture recognition system with an accuracy of 95% for a set of 5 gestures. Nguyen [40] described a hand gesture recognition system using a real-time tracking method with pseudo two-dimensional Hidden Markov Models. Chen [41] used it in combination with Fourier descriptors for hand gesture recognition using a real-time tracking method. Kelly [42] implemented an extension to the standard HMM model to develop a gesture threshold HMM (GT-HMM) framework which is specifically designed to identify inter gesture transition. Zafrulla [43] have investigated the potential of the Kinect [44, 45] depth-mapping camera for sign language recognition and verification for educational games for deaf children. They used 4-state HMMs to train each of the 19 signs defined in their study. Cooper [46] implemented an isolated sign recognition system using a 1st order Markov chain. In their model, signs are broken down in visemes (equivalent to phonemes in speech) and a bank of Markov chains are used to recognize the visemes as they are produced. Trigueiros used HMMs in a vision-based system able to interpret gestures from a robotic soccer referee [47], and also in a generic solution, able to interpret user commands, composed of a set of dynamic and static gestures, and that can be used to build any human-computer interaction (HCI) system [48]. His solution has the advantage of being computationally simple, easy to train and use, and at the same time generic enough, allowing its application in any HCI interface.

3 Approach

The Sign Language Recognition Prototype is a real-time vision-based system whose purpose is to recognize the Portuguese Sign Language given in the alphabet of **Fig. 1**. The purpose of the prototype was to test the validity of a vision-based system for sign language recognition and at the same time, test and select hand features that could be used with machine learning algorithms allowing their application in any real-time sign language recognition systems. For that, the user must be positioned in front of the camera, doing the sign language gestures, that will be interpreted by the system and their classification will be displayed on the right side of the interface. The implemented solution uses only one camera, a Kinect camera [44], and is based on a set of assumptions, hereby defined:

1. The user must be within a defined perimeter area, in front of the camera.
2. The user must be within a defined distance range, due to camera limitations. The system-defined values are 0.7m for the near plane and 3m for the far plane.
3. Hand pose is defined with a bare hand and not occluded by other objects.
4. The system must be used indoor, since the selected camera does not work well under sun light conditions.

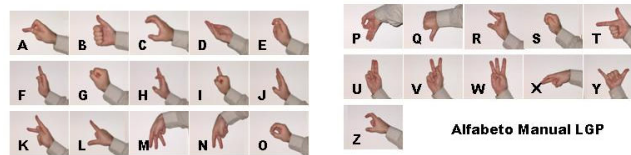


Fig. 1. Manual alphabet for the Portuguese Language.

The diagram of **Fig. 2** shows the proposed system architecture, which consists of two modules, namely: data acquisition, pre-processing and feature extraction; and sign language gesture classification.

In the first module, the hand is detected, tracked and segmented from the video images. From the obtained segmented hand, features are extracted for gesture classification. In the gesture classification module, the obtained feature vector (instance vector) is normalized and classified with a previous trained Support Vector Machine (SVM), which is a pattern recognition technique in the area of supervised machine learning, which works very well with high-dimensional data.

3.1 Prototype Implementation

The human-computer interface (HCI) for the prototype was developed using the C++ language, and the openFrameworks toolkit [49] with the OpenCV [50] and the OpenNI [51] addons, ofxOpenCv and ofxOpenNI respectively. OpenCV was used for some of the vision-based operations like extracting the hand blob contour, and OpenNI was responsible for the RGB and depth image acquisition. For model training and gesture classification the open source Dlib library was used, a general-purpose cross-platform C++ library capable of SVM multiclass classification [52]. The SVM

algorithm was selected for the final implementation, because in the experiments that were carried out with the selected features, it was able to achieve very high values of accuracy. Also, the resulting obtained model was compact and fast, able to be applied in applications with real-time classification demands. In the following two images it is possible to see the Sign Language Prototype with two vowels correctly classified and displayed on the right side of the user interface.

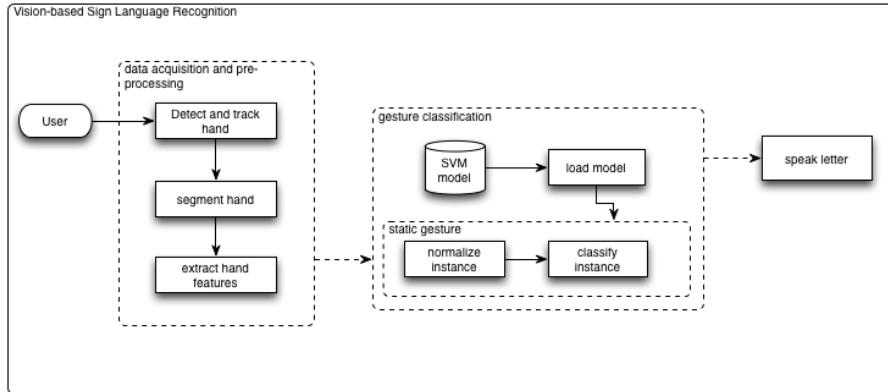


Fig. 2. Sign Language Recognition Prototype diagram

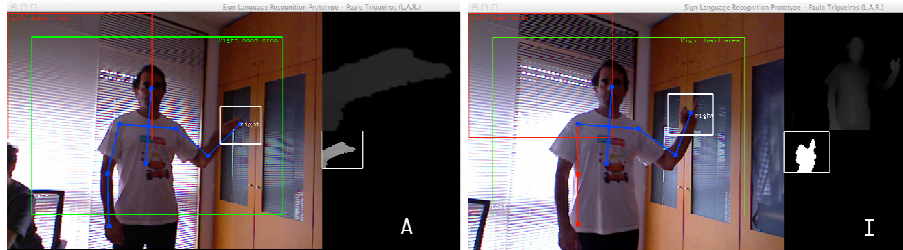


Fig. 3. Sign Language prototype interface wit two vowels correctly classified.

3.2 Experimental Methodology

This experiment main goal was to test two different types of hand features in order to validate which one could achieve better results in terms of Portuguese Sign Language recognition. The first type of features, the centroid distance features, derived from the object boundary coordinates are also called a shape signature as shown in [53, 54]. According to Zhang [55] and Trigueiros [8], it gives very good results in shape retrieval and classification, being at the same time simple in terms of computational complexity. The centroid distance is expressed by the distance of the hand contour boundary points $(x_i, y_i) : i = 0 \dots N-1$, from the centroid (x_c, y_c) of the hand and given by the following formula:

$$d(i) = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} : \quad i = 0, \dots, N \quad (1)$$

where $d(i)$ is the calculated distance. Due to the subtraction of centroid, which represents the hand position, from boundary coordinates, the centroid distance representation is invariant to translation. The second type of features is obtained from the hand depth image distance values. The distance values, which represent the distance from the object to the camera, can be seen in the image of **Fig. 4** (right), with different distances represented by different grey values. The distance values could represent, in our opinion, good hand features giving a more possible hand representations with the same viewpoint. For the centroid distance, the number of features (N) was defined to be 16. For the distance values, the hand image was resized to be 16x16 pixels in size, giving a final vector with 256 features. So, for this experiment, two types of images were used during the feature extraction phase and dataset construction: The binary hand blob for the centroid distance calculation (**Fig. 4** – left and centre) and The hand distance image for the distance feature vector (**Fig. 4** - right). For the centroid distance dataset (DataSet-1), a total of 2170 records with 16 features were used, and for the distance values dataset (DataSet-2), a total of 2488 records with 256 features were used.

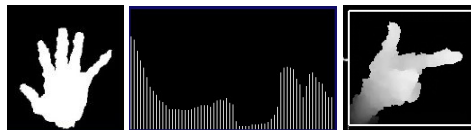


Fig. 4. Hand binary blob (left); Centroid distance signature (centre); Hand depth image (right)

All the built feature datasets were converted to Excel files, so that they can be imported into RapidMiner [4] for data analysis. The experiment was divided into two phases with the purpose of finding out for each dataset which parameters to use for classification. The first goal was to run a parameter optimization process, to find in terms of SVM classification, which kernels would achieve the best results among the following four: linear, sigmoid, rbf (radial basis function) and polynomial. Although the rbf kernel is a reasonable first choice, there are some situations where this is not suitable. In particular, when the number of features is very large, one may just use the linear kernel. This was the reason that led us to perform this test. The second goal was to run for each dataset a parameter optimization test dependent on the chosen kernel type, in order to find the best parameters to be implemented in the final solution.

3.3. Results

The first phase of the experiment, as explained in the previous section, was to learn which kernels behaved better in terms of SVM classification for the two datasets under study. As can be seen on

Table 1, for both datasets, the obtained best kernel type was the linear kernel, with a difference only on the obtained cost parameter C. For this reason the second part of the experiment was unnecessary, since the linear kernel does not need any further parameter optimization. Table 2 presents the obtained accuracy for each dataset with the corresponding kernel type and cost parameter.

Table 1. Parameters for each dataset.

Parameters	Dataset 1	Dataset 2
kernel type	linear	linear
C	1	2

Table 2. Accuracy for each dataset.

Parameters	Dataset 1	Dataset 2
Kernel type	Linear	linear
C	1	2
accuracy	99,4%	99,6%

In order to analyse how classification errors were distributed among classes, a confusion matrix (CM) was computed for each one of the datasets under study. The resulting confusion matrixes are represented in the following two tables.

Table 3. Centroid distance features CM

		Actual class				
		1	2	3	4	5
Predicted class	1	455	0	0	2	0
	2	0	394	1	1	0
	3	0	0	401	1	0
	4	4	2	0	382	0
	5	0	0	1	0	439

Table 4. Distance features CM

		Actual class				
		1	2	3	4	5
Predicted class	1	622	0	8	0	10
	2	0	543	1	0	1
	3	0	3	451	0	0
	4	0	0	0	413	0
	5	0	0	1	0	434

3.4 Discussion

The goal of this experiment was to carry a comparative study of two different types of hand features for the problem of sign language recognition using a Support Vector Machine (SVM). It was decided to choose this learning algorithm since it had already shown good results in previous experiments, in terms of classification, with the first type of hand features. From Table 2 one can easily see that the two different features gave similar results in terms of classification accuracy, with only a small difference of 0.2%. From the obtained results so far, one can conclude that the new features are not an asset, having the disadvantage of being heavier in terms of number of features and final sizes of the dataset and generated model. The corresponding confusion matrix also shows, that some major classification errors occurred between gesture one, the 'A' vowel, and gestures three and five, corresponding to the 'E' and 'U' vowels respectively. The centroid distance, on the other hand, being simple in terms of computational complexity gives rise to smaller datasets and a final small model file. Although the results were encouraging, further tests should be carried out, including not only more features from new users as well as the rest of the alphabet in the datasets in order to test the efficiency of the proposed system as well as the final classification model.

4 Conclusions and Future Work

Hand gestures are a powerful way for human communication, with lots of potential applications in the area of human computer interaction. Vision-based hand gesture recognition techniques have many proven advantages compared with traditional devices. However, hand gesture recognition is a difficult problem and the current work is only a small contribution towards achieving the results needed in the field of sign language recognition. This paper presented a vision-based system able to interpret static hand gestures from the Portuguese Sign Language. Experiments with two different datasets were carried out in order to find the best hand features, among two different types, in terms of Portuguese Sign Language gesture classification. The extracted features were tested with the help of the RapidMiner tool for machine learning and data mining. That way, it was possible to identify the best SVM parameters for learning and classification. The obtained parameters were able to achieve very good results in terms of real-time gesture classification with a minimal difference (0,2%) between the two hand features. The proposed solution was tested in real time situations, where it was possible to prove that the obtained classification models were able to recognize all the trained gestures being at the same time user independent, important requirements for this type of systems. The selected hand features, in conjunction with machine learning algorithms, proved to be very efficient, allowing their application in any real-time sign language recognition systems. As future work it is intended to keep improving the system and make experiments with complete language datasets. It is also intended to test systems able to interpret dynamic sign language gestures where there is a need for reliable solutions for the problem of start and end gesture identification. As a final conclusion one can say that although there is still much to do in the area, the proposed solution is a solid foundation for the development of any vision-based sign language recognition user interface system. The sign language grammar can be easily changed and the system configured to train the new language gestures.

Acknowledgments: The authors wish to thank all members of the Laboratório de Automação e Robótica (LAR), at University of Minho, Guimarães. We would like to thank IPP (Instituto Politécnico do Porto) for the contributions that made all this work possible. A special thanks also, to everyone that contributed to the project and made this work possible.

References

1. Wikipedia. *Língua gestual portuguesa*. 2012 September 9, 2013 [cited 2013; Available from: http://pt.wikipedia.org/wiki/Língua_gestual_portuguesa.
2. Vijay, P.K., et al., *Recent Developments in Sign Language Recognition : A Review*. International Journal on Advanced Computer Engineering and Communication Technology, 2012. **1**(2): p. 21-26.

3. Mingqiang, Y., K. Idiyo, and R. Joseph, *A Survey of Shape Feature Extraction Techniques*. Pattern Recognition, 2008: p. 43-90.
4. Miner, R. *RapidMiner : Report the Future*. December 2011]; Available from: <http://rapid-i.com/>.
5. Mitra, S. and T. Acharya, *Gesture recognition: A Survey*, in *IEEE Transactions on Systems, Man and Cybernetics*2007, IEEE. p. 311-324.
6. Murthy, G.R.S. and R.S. Jadon, *A Review of Vision Based Hand Gestures Recognition*. International Journal of Information Technology and Knowledge Management, 2009. 2(2): p. 405-410.
7. Wachs, J.P., H. Stern, and Y. Edan, *Cluster Labeling and Parameter Estimation for the Automated Setup of a Hand-Gesture Recognition System*. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 2005. 35(6): p. 932-944.
8. Trigueiros, P., F. Ribeiro, and L.P. Reis. *A Comparative Study of different image features for hand gesture machine learning*. in *5th International Conference on Agents and Artificial Intelligence*. 2013. Barcelona, Spain.
9. Trigueiros, P. and F. Ribeiro. *Vision-based Hand WheelChair Control*. in *12th International Conference on Autonomous Robot Systems and Competitions*. 2012. Guimarães, Portugal.
10. Wang, C.-C. and K.-C. Wang. *Hand Posture Recognition Using Adaboost with SIFT for Human Robot Interaction*. in *Proceedings of the International Conference on Advanced Robotics* 2008. Jeju, Korea.
11. Conseil, S., S. Bourenname, and L. Martin, *Comparison of Fourier Descriptors and Hu Moments for Hand Posture Recognition*, in *15th European Signal Processing Conference (EUSIPCO)2007*: Poznan, Poland. p. 1960-1964.
12. Barczak, A.L.C., et al., *Analysis of Feature Invariance and Discrimination for Hand Images: Fourier Descriptors versus Moment Invariants*, in *International Conference Image and Vision Computing*2011: New Zeland.
13. Bourennane, S. and C. Fossati, *Comparison of shape descriptors for hand posture recognition in video*. Signal, Image and Video Processing, 2010. 6(1): p. 147-157.
14. Triesch, J. and C.v.d. Malsburg. *Robust Classification of Hand Postures against Complex Backgrounds*. in *International Conference on Automatic Face and Gesture Recognition*. 1996. Killington, Vermont, USA.
15. Huynh, D.Q. *Evaluation of Three Local Descriptors on Low Resolution Images for Robot Navigation*. in *24th International*

- Conference Image and Vision Computing*. 2009. Wellington, New Zealand.
16. Fang, Y., et al. *Hand Posture Recognition with Co-Training*. in *19th International Conference on Pattern Recognition*. 2008. Tampa, FL, USA.
 17. Blum, A. and T. Mitchell, *Combining labeled and unlabeled data with co-training*, in *Proceedings of the eleventh annual conference on Computational learning theory*1998, ACM: Madison, Wisconsin, United States. p. 92-100.
 18. Tara, R.Y., P.I. Santosa, and T.B. Adj, *Sign Language Recognition in Robot Teleoperation using Centroid Distance Fourier Descriptors*. *International Journal of Computer Applications*, 2012. **48**(2).
 19. Faria, B.M., N. Lau, and L.P. Reis. *Classification of Facial Expressions Using Data Mining and machine Learning Algorithms*. in *4ª Conferência Ibérica de Sistemas e Tecnologias de Informação*. 2009. Póvoa de Varim, Portugal.
 20. Faria, B.M., L.P. Reis, and N. Lau. *Cerebral Palsy EEG Signals Classification: Facial Expressions and Thoughts for Driving an Intelligent Wheelchair*. in *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*. 2012.
 21. Gillian, N.E., *Gesture Recognition for Musician Computer Interaction*, in *Music Department*2011, Faculty of Arts, Humanities and Social Sciences: Belfast. p. 206.
 22. Faria, B.M., et al., *Machine Learning Algorithms applied to the Classification of Robotic Soccer Formations and Opponent Teams*, in *IEEE Conference on Cybernetics and Intelligent Systems (CIS)2010*: Singapore. p. 344 - 349
 23. Mannini, A. and A.M. Sabatini, *Machine learning methods for classifying human physical activity from on-body accelerometers*. *Sensors (Basel)*, 2010. **10**(2): p. 1154-75.
 24. Maldonado-Báscon, S., et al., *Road-Sign detection and Recognition Based on Support Vector Machines*, in *IEEE Transactions on Intelligent Transportation Systems*2007. p. 264-278.
 25. Vicen-Bueno, R., et al., *Complexity Reduction in Neural Networks Applied to Traffic Sign Recognition Tasks*, 2004.
 26. Trigueiros, P., F. Ribeiro, and L.P. Reis. *A comparison of machine learning algorithms applied to hand gesture recognition*. in *7th Iberian Conference on Information Systems and Technologies*. 2012. Madrid, Spain.
 27. Rego, P., P.M. Moreira, and L.P. Reis. *Serious games for rehabilitation: A survey and a classification towards a taxonomy*. in *Information Systems and Technologies (CISTI), 2010 5th Iberian Conference on*. 2010.

28. Rego, P.A., P.M. Moreira, and L.P. Reis. *Natural user interfaces in serious games for rehabilitation*. in *Information Systems and Technologies (CISTI), 2011 6th Iberian Conference on*. 2011.
29. Mendes, L., et al. *Virtual centre for the rehabilitation of road accident victims (VICERAVI)*. in *Information Systems and Technologies (CISTI), 2012 7th Iberian Conference on*. 2012.
30. Rego, P.A., P.M. Moreira, and L.P. Reis, *New Forms of Interaction in Serious Games for Rehabilitation*, in *Handbook of Research on Serious Games as Educational, Business and Research Tools*. 2012, IGI Global. p. 1188-1211.
31. Montesano, L., et al., *Towards an Intelligent Wheelchair System for Users With Cerebral Palsy*. *Neural Systems and Rehabilitation Engineering*, IEEE Transactions on, 2010. **18**(2): p. 193-202.
32. Braga, R.A., et al., *IntellWheels: modular development platform for intelligent wheelchairs*. *J Rehabil Res Dev*, 2011. **48**(9): p. 1061-76.
33. Faria, B.M., et al., *Evaluation of distinct input methods of an intelligent wheelchair in simulated and real environments: a performance and usability study*. *Assistive technology : the official journal of RESNA*, 2013. **25**(2): p. 88-98.
34. Faria, B.M., et al. *A Methodology for Creating Intelligent Wheelchair Users' Profiles*. in *4th International Conference on Agents and Artificial Intelligence*. 2012. Algarve, Portugal.
35. Ke, W., et al., *Real-Time Hand Gesture Recognition for Service Robot*. 2010: p. 976-979.
36. Oshita, M. and T. Matsunaga. *Automatic learning of gesture recognition model using SOM and SVM*. in *International Conference on Advances in Visual Computing*. 2010. Las Vegas, NV, USA: Springer-Verlag.
37. Almeida, R., L.P. Reis, and A.M. Jorge, *Analysis and Forecast of Team Formation in the Simulated Robotic Soccer Domain*, in *Proceedings of the 14th Portuguese Conference on Artificial Intelligence: Progress in Artificial Intelligence 2009*, Springer-Verlag: Aveiro, Portugal. p. 239-250.
38. Oka, K., Y. Sato, and H. Koike, *Real-time fingertip tracking and gesture recognition*. *IEEE Computer Graphics and Applications*, 2002. **22**(6): p. 64-71.
39. Perrin, S., A. Cassinelli, and M. Ishikawa. *Gesture recognition using laser-based tracking system*. in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*. 2004. Seoul, South Korea.
40. Binh, N.D., E. Shuichi, and T. Ejima. *Real-Time Hand Tracking and Gesture Recognition System*. in *Proceedings of International Conference on Graphics, Vision and Image 2005*. Cairo - Egypt.

41. Chen, F.-S., C.-M. Fu, and C.-L. Huang, *Hand gesture recognition using a real-time tracking method and hidden Markov models*. Image and Vision Computing, 2003. **21**(8): p. 745-758.
42. Kelly, D., J. McDonald, and C. Markham, *Recognition of Spatiotemporal Gestures in Sign Language Using Gesture Threshold HMMs*, in *Machine Learning for Vision-Based Motion Analysis*, L. Wang, et al., Editors. 2011, Springer London. p. 307-348.
43. Zafrulla, Z., et al., *American sign language recognition with the kinect*, in *13th International Conference on Multimodal Interfaces2011*, ACM: Alicante, Spain. p. 279-286.
44. Chowdhury, J.R., *Kinect Sensor for Xbox Gaming*, 2012, IIT Kharagpur.
45. Andersen, M.R., et al., *Kinect Depth Sensor Evaluation for Computer Vision Applications*, in *Technical report ECE-TR-6 2012*, Department of Engineering – Electrical and Computer Engineering, Aarhus University. p. 37.
46. Helen, C. and B. Richard. *Large lexicon detection of sign language*. in *11th International Conference on Human-Computer Interaction*. 2007. Rio de Janeiro, Brazil: Springer-Verlag.
47. Trigueiros, P., F. Ribeiro, and L.P. Reis. *Vision Based Referee Sign Language Recognition System for the RoboCup MSL League*. in *17th annual RoboCup International Symposium*. 2013. Eindhoven, Holland.
48. Trigueiros, P., F. Ribeiro, and L.P. Reis. *Vision-based Gesture Recognition System for Human-Computer Interaction*. in *IV ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing*. 2013. Funchal. Madeira: Taylor and Francis, Publication
49. Lieberman, Z., T. Watson, and A. Castro. *openFrameworks*. 2004 10 October 2013 [cited 2011; openFrameworks is an open source C++ toolkit designed to assist the creative process by providing a simple and intuitive framework for experimentation]. Available from: <http://www.openframeworks.cc/>.
50. Bradski, G. and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*. 1st ed. 2008: O'Reilly Media.
51. OpenNI. *The standard framework for 3D sensing*. 2013; Available from: <http://www.openni.org/>.
52. King, D.E., *Dlib-ml: A Machine Learning Toolkit*. Journal of Machine Learning Research, 2009. **10**: p. 1755-1758.
53. Zhang, D. and G. Lu. *A comparative Study of Fourier Descriptors for Shape Representation and Retrieval*. in *Proc. of 5th Asian Conference on Computer Vision (ACCV)*. 2002. Melbourne, Australia: Springer.

54. Kauppinen, H., T. Seppanen, and M. Pietikainen, *An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995. **17**(2): p. 201-207.
55. Zhang, D. and G. Lu, *A Comparative Study on Shape Retrieval Using Fourier Descriptors with Different Shape Signatures*. Journal of Visual Communication and Image Representation, 2003(14 (1)): p. 41-60.