Generic System for Human-Computer Gesture Interaction

Paulo Trigueiros
ISCAP/IPP - Instituto Politécnico do Porto, Departamento de Informática, Centro ALGORITMI e LIACC - Lab. Inteligência Artificial e Ciência de Computadores
4200 - 465 Porto, Portugal
pjt@iscap.ipp.pt

Fernando Ribeiro
EEUM/DEI - Escola de Engenharia da Universidade do Minho, Departamento de Eletrónica Industrial,
Centro ALGORITMI
Campus de Azurém 4800-058, Guimarães, Portugal
fernando@dei.uminho.pt

Luis Paulo Reis
EEUM/DSI - Escola de Eng. da Univ. do Minho, Dep. Sistemas de Informação, LIACC – Lab. Inteligência Artificial e Ciência de Computadores e Centro ALGORITMI
Campus de Azurém 4800-058, Guimarães, Portugal
lpreis@dsi.uminho.pt

***Abstract*— 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, giving users a simpler and more natural way to communicate with electronic devices. This work proposes a generic system architecture based in computer vision and machine learning, able to be used with any interface for human-computer interaction. The proposed solution is mainly composed of three modules: a pre-processing and hand segmentation module, a static gesture interface module and a dynamic gesture interface module. The experiments showed that the core of vision-based interaction systems can be the same for all applications and thus facilitate the implementation. In order to test the proposed solutions, three prototypes were implemented. For hand posture recognition, a SVM model was trained and used, able to achieve a final accuracy of 99.4%. For dynamic gestures, an HMM model was trained for each gesture that the system could recognize with a final average accuracy of 93.7%. The proposed solution as the advantage of being generic enough with the trained models able to work in real-time, allowing its application in a wide range of human-machine applications.**

Keywords— Human-computer interaction; Gesture interfaces; Generic systems; Computer Vision; Machine Learning

I. INTRODUCTION

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, giving users a simpler and more intuitive way of communication between a human and a computer. Using visual input in this context makes it possible to communicate remotely with computerized equipment, without the need for physical contact. For gesture-based applications, we need to model them 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. 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, etc. There are areas where this trend is an asset, as for example in the application of these technologies in interfaces that can help people with physical disabilities, or areas where it is a complement to the normal way of communicating.

The main objective of this work is to describe and propose a generic system architecture based in computer vision and machine learning, able to be used with any interface for human-machine interaction

The rest of the paper is as follows. Firstly, the related work is review in section 2. Section 3 introduces the system architecture and describes each one of the proposed modules. Experimental methodology and results are explained in section 4. Conclusions are drawn in section 5.

II. RELATED WORK

Hand gestures, either static or dynamic, for human computer interaction in real time systems is an area of active research and with many possible applications. However, vision-based hand gesture interfaces for real-time applications require fast and extremely robust hand detection, feature extraction and gesture recognition. Several approaches are normally used including Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Hidden Markov Models (HMM).

An Artificial Neural Networks is a mathematical / computational model that attempts to simulate the structure of biological neural systems. They accept features as inputs and produce decisions as outputs [1]. Maung et al [2] applied it in a gesture recognition system for real-time gestures in unstrained environments. Vicen-Buéno et al. [3] used it applied to the problem of traffic sign recognition. Bailador et al. [4] presented an approach to the problem of gesture recognition in real time using inexpensive accelerometers. Their approach was based on the idea of creating specialized signal predictors for each gesture class.

A Support Vector Machines (SVM's) is a technique based on statistical learning theory, which works very well with high-dimensional data. The objective of this algorithm is to find the optimal separating hyper plane between two classes by maximizing the margin between them [5]. Faria [6, 7] used it to classify robotic soccer formations and the classification of facial expressions, Ke [8] used it in the implementation of a real-time hand gesture recognition system for human robot interaction, Maldonado-Báscon [9] used it for the recognition of road-signs and Masaki [10] used it in conjunction with 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 [11] 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.

Trigueiros [12] has 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. He has also made a comparative study of different image features for hand gesture machine learning [13] and proposed a vision-based system for the Portuguese sign language recognition [14], based in a SVM model with an accuracy of 99.2%, and a Vision-based Gesture Recognition System for Human Computer Interaction [15] based in machine learning algorithms and able to do real-time hand gesture recognition.

Hidden Markov Models (HMMs) have been widely used in a successfully way in speech recognition and hand writing recognition [16], in various fields of engineering and also applied quite successfully to gesture recognition.

Oka [17] 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 [18] 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 [19] described a hand gesture recognition system using a real-time tracking method with pseudo two-dimensional Hidden Markov Models. Chen [20] used it in combination with Fourier descriptors for hand gesture recognition using a real-time tracking method. Kelly [21] 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 [22] have investigated the potential of the Kinect 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. Trigueiros [23] used HMM's applied to a Vision-based system capable of recognizing a set of referee commands for robotic soccer games. Cooper [24] 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. Milosevic [25] implemented an HMM-based continuous gesture recognition algorithm, optimized for lower power, low cost microcontrollers without float point unit. The proposed solution is validated on a set of gestures performed with the Smart Micrel Cube (SMCube), which embeds a 3-axis accelerometer and an 8-bit microcontroller. They also explore a multiuser scenario where up to 4 people share the same device. Elmezain [26] proposed a system able to recognize both isolated and continuous gestures for Arabic numbers (0-9) in real-time. To handle isolated gestures, an HMM using Ergodic (it is possible to go from every state to every state), Left-Right (LR) and Left-Right Banded (LRB) topologies with different number of states was applied. The LRB in conjunction with the Forward algorithm presented the best performance with an average recognition rate of 98.94% and 95.7% for isolated and continuous gestures.

III. Proposed System Architecture

The design of any gesture recognition system essentially involves the following three aspects: (1) *data acquisition and pre-processing*; (2) *data representation or feature extraction* and (3) *classification or decision-making*. Taking this into account, a possible solution to be used in any vision-based hand gesture recognition system for human-machine interaction is represented in the diagram of Fig. 1.

In the following sections, we will describe the *Static gesture module* and the *Dynamic gesture module.*

A. Static Gesture Module

For static gesture classification, hand segmentation and feature extraction is a crucial step in vision-based hand gesture recognition systems. The obtained segmented hand in the pre-processing module is used to extract hand features that are used later with classification algorithms [13]. The learned models for hand posture classification use feature vectors composed of centroid distance values. The centroid distance signature is a type of shape signature [27] expressed by the distance of the hand contour boundary points, from the hand centroid (xc, yc) and is calculated in the following manner:

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

This way, a one-dimensional function representing the hand shape is obtained. The number of equally spaced points N used in our implementation was 16. The feature vectors thus obtained were used to train the SVM used in system implementations. The SVM is a pattern recognition technique in the area of supervised machine learning, which works very well with high-dimensional data. When more than two classes are present, there are several approaches that evolve around the 2-class case [28]. The one used in this system is the one-against-all, where c classifiers have to be designed. Each one of them is designed to separate one class from the rest.

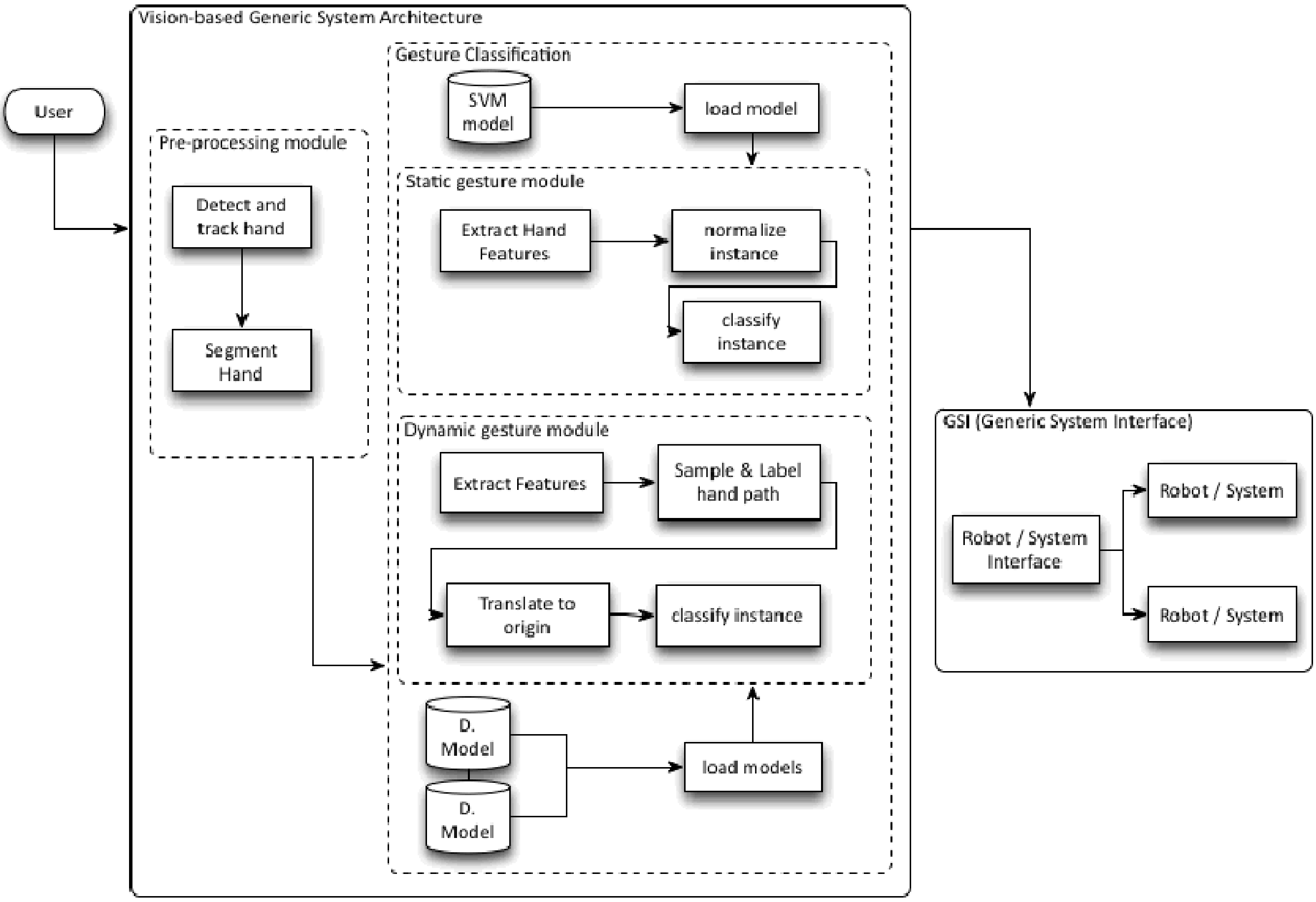

Fig. 1. Generic system architecture for a human-computer gesture interface

B. Dynamic gesture module

Dynamic gestures are time-varying processes, which show statistical variations, making HMMs a plausible choice for modelling the processes [29] [30]. So, for the recognition of dynamic gestures a HMM (Hidden Markov Model) model was trained for each possible gesture. A Markov Model is a typical model for a stochastic (i.e. random) sequence of a finite number of states [31].

A human gesture can be understood as a Hidden Markov Model where the true states of the model are hidden in the sense that they cannot be directly observed. HMMs have been widely used in a successfully way in speech recognition and hand writing recognition [16]. In this system the 2D motion hand trajectory points are labelled according to the distance to the nearest centroid, based on Euclidean distance, and translated to origin resulting in a discrete feature vector. The feature vectors thus obtained are used to train the different HMMs and learn the model parameters: the initial state probability vector (Π), the state-transition probability matrix ($A=[a_{ij}]$) and the observable symbol probability matrix ($B=[b_j(m)]$). In the recognition phase an output score for the sample gesture is calculated for each model, given the likelihood that the corresponding model generated the underlying gesture. The model with the highest output score represents the recognized gesture. In our system a Left-Right (LR) HMM, like the one shown in Fig. 2, was used [32, 33]. This kind of HMM has the states ordered in time so that as time increases, the state index increases or stays the same. This topology has been chosen, since it is perfectly suitable to model the kind of temporal gestures present in the system.

IV. Experimental Methodology and Results

The experimental methodology was divided into three parts: a hand posture database creation with the selected features and SVM model training, a dynamic gesture database creation and HMM model training and an implementation of three prototypes, in order to test the proposed architecture and able to do hand posture recognition, integration of static and dynamic gesture recognition and sign language recognition.

For hand posture recognition an SVM model was trained based on a dataset build from data collected from four users making five pre-defined postures in front of a Kinect camera. In order to analyse the best parameters for SVM, the extracted features were analysed with the help of Rapid Miner [34].

The experiments were performed with parameter optimization for the cost parameter C, with a 10-fold cross validation, and we obtained an accuracy of 99.4% with a linear kernel with a C value equal to 1.

TABLE I. Hidden Markov Models accuracy for each gesture trained

Gesture	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Accuracy	75%	100%	100%	100%	92%	88%	92%	100%	100%	96%	88%

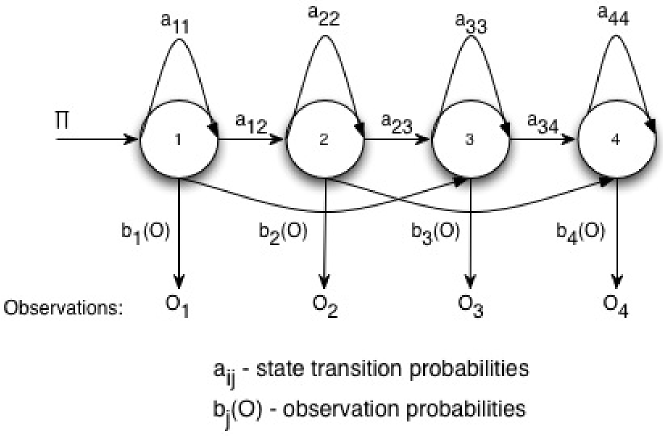
a11
a22
a33
a44
π
a12
a23
a34
1
2
3
4
b1(O)
b2(O)
b3(O)
b4(O)
Observations:
O1
O2
O3
O4
aij - state transition probabilities
bj(O) - observation probabilities

Fig. 2. A 4-state Left-Right HMM model.

TABLE II. Centroid distance features confusion matrix

		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**

The obtained confusion matrix is shown in TABLE II. For model training, the Dlib [35] library was used. This is a general-purpose cross-platform C++ library capable of SVM multiclass classification. For dynamic gesture recognition, an HMM model was trained off-line for each one of the 11 predefined gestures and the three parameters (the initial state probability vector, the state-transition probability matrix and the observable symbol probability matrix) were learned and saved. Once again four users were used to perform the predefined gestures and the extracted features were saved and used to train the models. The number of observation symbols (alphabet) and hidden states were learned by trial and error, and were defined to be 64 and 4 respectively. For model training and implementation it was decided to use an openFrameworks [36] add-on implementation of the HMM algorithm for classification and recognition of numeric sequences. This implementation is a C++ porting of a MATLAB code from Kevin Murphy [37].

The test datasets obtained were analysed with the learned models and the final accuracy results obtained are represented in TABLE I.

So, for the dynamic gesture recognition an average accuracy of 93.7% was achieved.

The first prototype (Fig. 3) is an application able to recognize in real-time a set of hand postures that can be used as commands to control any electronic device in a remote way. For that, the system uses the trained models to classify the hand feature vector as can be seen in the figure. The second one, is an application able to build a sequence of static and dynamic gestures that can be used also as a command in any human-machine interaction system as shown in Fig. 4. The final one, shown in Fig. 5, is a real-time system able to do sign language recognition.

V. Conclusions

This paper presented a system able to interpret dynamic and static gestures from a user with the goal of real-time human-computer interaction. Although the machine learning algorithms used are not the only solutions, they were selected based on obtained performance with the selected features. Thus, for hand posture classification a SVM model was learned from centroid distance features and a recognition rate of 99.4% was achieved. For dynamic gesture classification, a HMM model was learned for each gesture and a final average accuracy of 93.7% was achieved. We were able to test the system in real time situations, and it was possible to prove from the experiments that the trained models were able to recognize all the trained gestures, proving that this kind of models, as was already seen in other references, works very well for this type of problem. The experimental results also showed, that the proposed system was able to reliably recognize the pre-defined commands.

Prototype implementation was able to prove that the core of vision-based interaction systems can be the same for all application, and that the proposed generic system architecture is a solid foundation for the development of hand gesture recognition systems that can be integrated in any human-machine interface application.

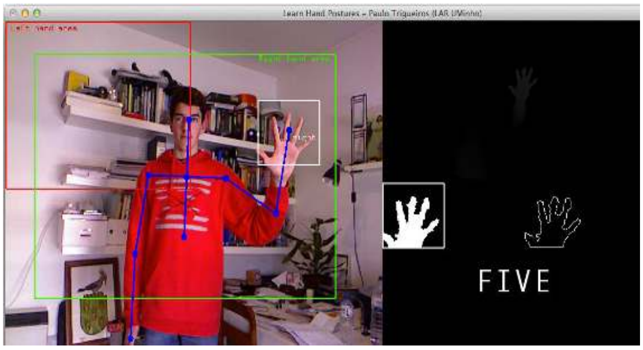

Fig. 3. Hand posture recognition prototype interface.

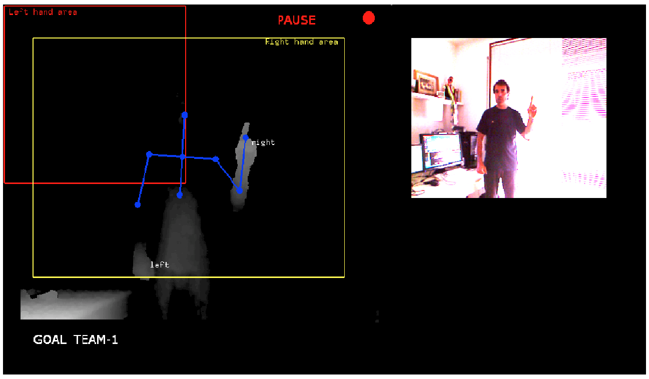

Fig. 4. Static and dynamic gesture recognition prototype interface.

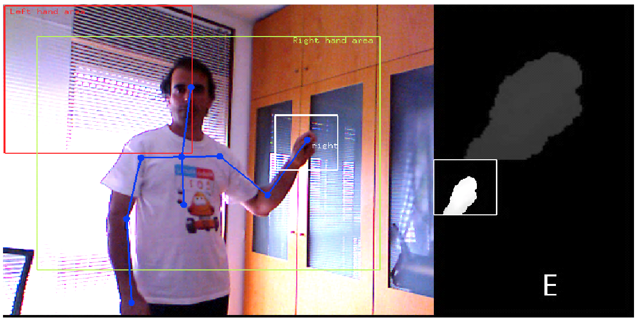

Fig. 5. Sign language recognition prototype interface.

REFERENCES

[1] W. E. Snyder and H. Qi, Machine Vision: Cambridge University Press, 2004.

[2] T. H. H. Maung, "Real-Time Hand Tracking and Gesture Recognition System Using Neural Networks," Proceedings of World Academy of Science: Engineering & Technology, vol. 50, pp. 466-470, 2009.

[3] R. Vicen-Bueno, R. Gil-Pita, M. P. Jarabo-Amores, and F. López-Ferreras, "Complexity Reduction in Neural Networks Appplied to Traffic Sign Recognition Tasks," 2004.

[4] G. Bailador, D. Roggen, G. Tröster, and G. Triviño, "Real time gesture recognition using continuous time recurrent neural networks," in 2nd International Conference on Body Area Networks, Florence, Italy, 2007, pp. 1-8.

[5] A. Ben-Hur and J. Weston, "A User's Guide to Support Vector Machines," in Data Mining Techniques for the Life Sciences. vol. 609, ed: Humana Press, 2008, pp. 223-239.

[6] B. M. Faria, 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, Póvoa de Varim, Portugal, 2009, pp. 197-206.

[7] B. M. Faria, L. P. Reis, N. Lau, and G. Castillo, "Machine Learning Algorithms applied to the Classification of Robotic Soccer Formations and Opponent Teams," presented at the IEEE Conference on Cybernetics and Intelligent Systems (CIS), Singapore, 2010.

[8] W. Ke, W. Li, L. Ruifeng, and Z. Lijun, "Real-Time Hand Gesture Recognition for Service Robot," pp. 976-979, 2010.

[9] S. Maldonado-Báscon, S. Lafuente-Arroyo, P. Gil-Jiménez, and H. Gómez-Moreno. (2007, June 2007) Road-Sign detection and Recognition Based on Support Vector Machines. IEEE Transactions on Intelligent Transportation Systems. 264-278. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4220659

[10] M. Oshita and T. Matsunaga, "Automatic learning of gesture recognition model using SOM and SVM," in International Conference on Advances in Visual Computing, Las Vegas, NV, USA, 2010, pp. 751-759.

[11] R. Almeida, L. P. Reis, and A. M. Jorge, "Analysis and Forecast of Team Formation in the Simulated Robotic Soccer Domain," presented at the Proceedings of the 14th Portuguese Conference on Artificial Intelligence: Progress in Artificial Intelligence, Aveiro, Portugal, 2009.

[12] P. Trigueiros, 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, Madrid, Spain, 2012, pp. 41-46.

[13] P. Trigueiros, 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, Barcelona, Spain, 2013.

[14] P. Trigueiros, F. Ribeiro, and L. P. Reis, "Vision-based Sign Language Recognition System," in World Conference on Information Systems and Technologies Madeira, Portugal, 2014.

[15] P. Trigueiros, 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, Funchal. Madeira, 2013.

[16] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE, vol. 77, pp. 257-286, 1989.

[17] K. Oka, Y. Sato, and H. Koike, "Real-time fingertip tracking and gesture recognition," IEEE Computer Graphics and Applications, vol. 22, pp. 64-71, 2002.

[18] S. Perrin, A. Cassinelli, and M. Ishikawa, "Gesture recognition using laser-based tracking system," in Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, South Korea, 2004, pp. 541-546.

[19] N. D. Binh, E. Shuichi, and T. Ejima, "Real-Time Hand Tracking and Gesture Recognition System," in Proceedings of International Conference on Graphics, Vision and Image Cairo - Egypt, 2005, pp. 362--368.

[20] F.-S. Chen, C.-M. Fu, and C.-L. Huang, "Hand gesture recognition using a real-time tracking method and hidden Markov models," Image and Vision Computing, vol. 21, pp. 745-758, 2003.

[21] D. Kelly, 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, G. Zhao, L. Cheng, and M. Pietik√§inen, Eds., ed: Springer London, 2011, pp. 307-348.

[22] Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P. Presti, "American sign language recognition with the kinect," presented at the 13th International Conference on Multimodal Interfaces, Alicante, Spain, 2011.

[23] P. Trigueiros, F. Ribeiro, and L. P. Reis, "Vision Based Referee Sign Language Recognition System for the RoboCup MSL League," in 17th annual RoboCup International Symposium, Eindhoven, Holland, 2013.

[24] C. Helen and B. Richard, "Large lexicon detection of sign language," in 11th International Conference on Human-Computer Interaction, Rio de Janeiro, Brazil, 2007, pp. 88-97.

[25] B. Milosevic, E. Farella, and L. Benini, "Continuous Gesture Recognition for Resource Constrained Smart Objects," in The Fourth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, Florence, Italy, 2010, pp. 391-396.

[26] M. Elmezain, A. Al-Hamadi, J. Appenrodt, and B. Michaelis, "A Hidden Markov Model-based Continuous Gesture Recognition System for Hand Motion Trajectory," in 19th International Conference on Pattern Recognition, Tampa, Florida, USA, 2008, pp. 1-4.

[27] P. Trigueiros, "Hand Gesture Recognition System based in Computer Vision and Machine Learning: Applications on Human-Machine Interaction," PhD, Electronics and Computer Engineering, University of Minho, 2014.

[28] S. Theodoridis and K. Koutroumbas, An Introduction to Pattern Recognition: A Matlab Approach: Academic Press, 2010.

[29] L. R. Rabiner and B. H. Juang, "An introduction to hidden Markov models," IEEE ASSp Magazine, 1986.

[30] Y. Wu and T. S. Huang, "Vision-Based Gesture Recognition: A Review," presented at the Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction, 1999.

[31] G. A. Fink, Markov Models for Pattern recognition - From Theory to Applications: Springer, 2008.

[32] F. Camastra and A. Vinciarelli, Machine Learning for Audio, Image and Video Analysis: Springer, 2008.

[33] E. Alpaydin, Introduction to Machine Learning: MIT Press, 2004.

[34] R. Miner. (December 2011). RapidMiner : Report the Future. Available: http://rapid-i.com/

[35] D. E. King, "Dlib-ml: A Machine Learning Toolkit," Journal of Machine Learning Research, vol. 10, pp. 1755-1758, 2009.

[36] Z. Lieberman, T. Watson, and A. Castro. (2004). openFrameworks. Available: http://www.openframeworks.cc/

[37] K. Murphy. (1998). Hidden Markov Model (HMM) Toolbox for Matlab. Available: Hidden Markov Model (HMM) Toolbox for Matlab