# Online Robot Teleoperation Using Human Hand Gestures: A Case Study for Assembly Operation

**Nuno Mendes, Pedro Neto, Mohammad Safeea and António Paulo Moreira**

**Abstract** A solution for intuitive robot command and fast robot programming is presented to assemble pins in car doors. Static and dynamic gestures are used to instruct an industrial robot in the execution of the assembly task. An artificial neural network (ANN) was used in the recognition of twelve static gestures and a hidden Markov model (HMM) architecture was used in the recognition of ten dynamic gestures. Results of these two architectures are compared with results displayed by a third architecture based on support vector machine (SVM). Results show recognition rates of 96 % and 94 % for static and dynamic gestures when the ANN and HMM architectures are used, respectively. The SVM architecture presents better results achieving recognition rates of 97 % and 96 % for static and dynamic gestures, respectively.

**Keywords** Gesture spotting · Robot programming · Robotic assembly · Industrial robot

## 1 Introduction

Robot programming is a time consuming and monotonous task. In order to speed up this task several approaches have been proposed, the majority of them work similarly to a traditional joystick and make use of accelerometers, electromyography and vision systems, besides that CAD-based systems have been extensible

N. Mendes(✉) · A.P. Moreira
Centre for Robotics and Intelligent Systems,
Institute for Systems and Computer Engineering Technology and Science, Porto, Portugal
e-mail: nuno.m.mendes@inesctec.pt, amoreira@fe.up.pt

P. Neto · M. Safeea
Centre for Mechanical Engineering of the University of Coimbra,
University of Coimbra, Coimbra, Portugal
e-mail: pedro.neto@dem.uc.pt, safeea@student.dem.uc.pt

L.P. Reis et al. (eds.), *Robot 2015: Second Iberian Robotics Conference*,
Advances in Intelligent Systems and Computing 418,
DOI: 10.1007/978-3-319-27149-1_8

explored. The big problem in these systems is the fact that they are limited to perform robot movements, or its user has to carry uncomfortable devices, and in some cases they are not intuitive. In order to make robots more user friendly and expand its use in industry, the way that a user interacts with a robot needs to be intuitive and no complex technical knowledge must be required.

The purpose of this study is the development of an intuitive robot programming system based on performed gestures. In order to achieve a feasible solution able to identify gestures performed by users, a vision system is used and gestures performed by a user are recognized by three different methods. The effectiveness of the proposed system is assessed through practical experiments identifying the recognition method that presents the best performance. Additionally, the system is tested in an industrial task of pin assembly in car doors.

## 2 State of the Art

Vision technology has been used in the recognition of human hands, face and body behaviors [1,2,3]. A major advantage of this technology is the fact that a user does not have to carry any device during the interaction process making the process more natural. In the last developments neither a mark is needed to be held by the user. A necessary requirement is the human presence in the vision sensor's field of view. Vision systems have difficulty producing robust information when facing cluttered environments. Some vision-based systems are view dependent, require a uniform background and illumination, and a single person (full-body or part of the body) in the camera field of view. In addition, occlusion of some reference points can occur frequently which need to be coped with.

There have been an increasing interest for gesture recognition using vision-based interfaces, for hand, arm and full-body gesture recognition [4]. Vision-based solutions have been used for real-time gesture spotting applied to the robotics field. A recent study presents a motion tracking system combining vision, inertial and magnetic sensing for spatial robot programming using gestures [5]. The main limitation of this approach is related to complex path execution. An American sign language (SL) word recognition system, that uses as interaction devices both a data glove and a motion tracker system, is presented in [6]. Inertial sensors have also been explored for different gesture-based applications [7], [8].

A number of machine learning techniques have been proposed to deal with gesture recognition, the most explored techniques rely on artificial neural network (ANN), hidden Markov models (HMM) and support vector machine (SVM). Mitra and Acharya provide a complete overview of techniques for gesture pattern recognition [9]. ANN-based problem solving techniques have been demonstrated to be a reliable tool in gesture recognition, presenting good learning and generalization capabilities. ANNs have been applied in a wide range of situations such as the recognition of continuous hand postures from gray-level video images, gesture recognition having acceleration data as input [10], full-body motion recognition for robot teleoperation and SL recognition [6]. The capacity of recurrent neural networks (RNNs) for modeling temporal sequence learning has also been demonstrated [11]. Nevertheless, RNNs are still difficult to train. HMMs are stochastic methods known for their application in temporal pattern recognition, including gesture spotting [12]. A survey on human-computer interaction using gesture recognition and vision-based systems as interaction technology was presented in [13]. It is concluded that research efforts are required to reliably recognize gestures in continuous and in relatively large libraries of gestures.

In order to recognize dynamic gestures different approaches have been employed, for example using discrete HMM to recognize online dynamic human hand gestures [14] and using HMM for full body gesture recognition [15]. Recognitions rates (RR) above 84 % were reported for a collection of seven dynamic gestures. A real-time system based on HMM was proposed by Kurakin et al. [16] for dynamic hand gesture recognition. This system takes into account variations in speed and human behavior as well as in hand orientations. A recognition rate of at least 76 % was achieved with this approach. Automatic recognition of facial emotion based on feed forward ANN and support vector regressors was presented by Zhang et al. [17]. An interesting study in the field reports a neural-based classifier for gesture recognition with an accuracy of over 99% for a library of only 6 gestures [18]. Other authors concluded that a hand contour-based neural network training is faster than complex moment-based neural network training but in the other hand the former proved to be less accurate (71%) than the latter (86%) [19].

## 3 Proposed Approach

A gesture recognition system is proposed in this study to command and program an industrial robot. This gesture recognition system relies on an infra-red stereo vision system that provides features about the human hands of a user to an intelligent recognition architecture. Three recognition architectures are approached and their results are compared among them. One of the recognition architectures based on ANN is used to recognize static gestures. On the other hand dynamic gestures are recognized in a second architecture based on HMM. Finally, a third architecture based on SVM is used to recognize the same static and dynamic gestures.

A Leap Motion Controller (LMC) provides a set of features $x$ , also called a data frame, about human hands and wrists. In this study some of these features were chosen to be used in recognition of human hand gestures. The features used are presented in Table 1 for static gestures and in Table 2 for dynamic gestures. Each set of features is acquired at a rate of 40 Hz. The static gesture recognition is carried out at 40 Hz and the dynamic gesture recognition is carried out at 3 Hz. The frame acquisition rate for recognition of dynamic gestures is lower because thirteen sets of features are required to perform the recognition of the dynamic gesture while just one set of features is required to perform the recognition of a static gesture. Different features are used for static and dynamic gestures, this is because:

- Number of gestures in the considered libraries;
- Number of differentiation variables required;
- Methods take a lot of processing time when the number of variables is higher;
- These gestures are better characterized/represented by different features.

**Table 1** Features of static gestures

| Data | Description |
|---|---|
| $x_1, x_2, x_3$ | x, y and z components of the normal vector to the palm hand, respectively |
| $x_4, x_5, x_6$ | x, y and z components of the palm hand direction vector, respectively |
| $x_7, x_8, x_9$ | x, y and z components of the thumb finger direction vector, respectively |
| $x_{10}, x_{11}, x_{12}$ | x, y and z components of the index finger direction vector, respectively |
| $x_{13}, x_{14}, x_{15}$ | x, y and z components of the middle finger direction vector, respectively |
| $x_{16}, x_{17}, x_{18}$ | x, y and z components of the ring finger direction vector, respectively |
| $x_{19}, x_{20}, x_{21}$ | x, y and z components of the pinkie finger direction vector, respectively |

**Table 2** Features of dynamic gestures

| Data | Description |
|---|---|
| $x_1, x_2, x_3$ | x, y and z components of the normal vector to the palm hand, respectively |
| $x_4$ | grab property |
| $x_5$ | x component of the thumb finger direction vector |

## 3.1 Artificial Neural Network

The ANN architecture used in this study consists of 21 input neurons, a hidden layer with 21 neurons and in the output layer is used 12 neurons. Table 3 presents a detailed parametrization of the ANN which is shown in Fig. 1. $\mathbf{W}$ represents a weight matrix, $\mathbf{b}$ is a weight vector, $\varphi$ represents an activation function and $y$ is the output of the ANN.

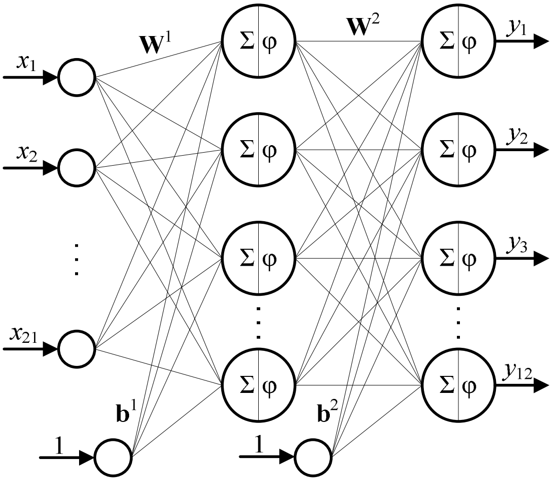

**Fig. 1** ANN architecture used in the recognition of static gestures

**Table 3** ANN parameters

| Parameter | Value |
|---|---|
| Number of neurons in the input layer | 21 |
| Number of neurons in the hidden layer | 21 |
| Number of neurons in the output layer | 12 |
| Activation function in the hidden layer | asymmetric sigmoid function, $\sigma = 1$ |
| Activation function in the output layer | asymmetric sigmoid function, $\sigma = 1$ |
| Learning coefficient | 0.25 |
| Momentum | 0.1 |
| Number of training cycles | 10 000 |
| Updating rate (ms) | 25 |

## 3.2 Hidden Markov Model

To recognize dynamic gestures the architecture based on HMMs shown in Fig. 2 is used. $s$ represents the states of the HMM and $a$ is the state transition probability. Table 4 presents the parametrization used in this architecture. A HMM based on left-right model is used for each dynamic gesture.

**Table 4** HMMs parameters

| Parameter | Value |
|---|---|
| Number of HMM | 10 |
| Number of states | 20 |
| Number of symbols | 3 |
| Stopping criterion - tolerance | 0.01 |

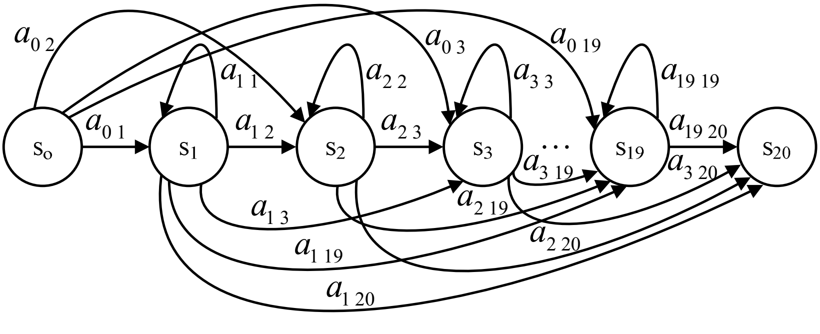

**Fig. 2** Left-right model of a HMM

## 3.3 Support Vector Machine

An algorithm based on SVM was trained to recognize static and dynamic gestures. The number of SVMs used in this system is the same as the number of predefined gestures, i.e. 12 SVMs for static gestures and 10 SVMs for dynamic gestures. A comparative analysis among the outputs of the different SVMs is done.

In order to reduce the size of the data and at the same time preserve its information, a transformation of the coordinate system was implemented. The data were normalized and transposed from Euclidian coordinate system to spherical coordinate system resulting in unit vectors represented only by two angles.

SVM is inherently a binary classifier that provides always an output as validated. However, the validated output can be wrongly classified and thus leading to false validations. In order to overcome these false classifications, the one versus one method was used and extended to deal with multi-class classification problems.

$n$ SVM algorithms are trained for $n$ gestures, each algorithm is taught to distinguish one of the gestures apart from the rest. During the prediction the input feature vector feeds all of the $n$ pre-trained SVMs and each one provides an outcome. The output of each SVM algorithm can assume one of the three following cases:

- A positive outcome from only one of the $n$ pre-trained SVM algorithms. In this case the gesture associated with that algorithm is assumed as recognized;
- Two or more positive predictions from the $n$ algorithms. In this case the result is assumed as uncertain considering the gesture as unrecognized;
- Negative predictions from all of the $n$ algorithms, then the gesture is considered as unrecognized.

## 4 Experiments

In order to test and assess the feasibility of the different proposed architectures, they are used in a practical experiment that consists in performing each gesture an hundred times and estimate its recognition rate. Figures 3 and 4 illustrate the static and dynamic gestures, respectively, used in this experiment.

Additionally, each gesture was associated to a robot command or function as can be seen in Figs. 3 and 4. The system was used to generate of robot code for an industrial task of pin assembly in car doors. In a first stage a data file is generated on-line with all information required to program an industrial robot, such as position, orientation and speed. After that, in a second stage, the previous data is prosprocessed and the robot code is generated.

In order to develop the gesture recognition system, execute it and perform the tests, a personal computer with the characteristics shown in Table 5 was used as well as an industrial robot ABB IRB140 equipped with IRC5 controller, a Leap Motion Controller and mechanical tools for the assembly task.

**Table 5** Personal computer characteristics used in this study

| Processor | Intel® Core$^{TM}$ i7-4700HQ CPU @ 2.40 GHz (8CPUs) |
|---|---|
| Memory | 8GB |
| Operating System | Microsoft Windows 8.1 64 bits |

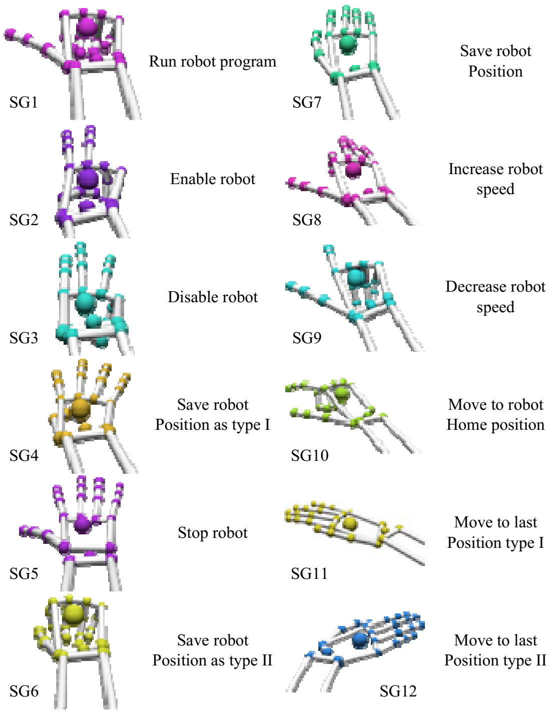

**Fig. 3** Static gestures

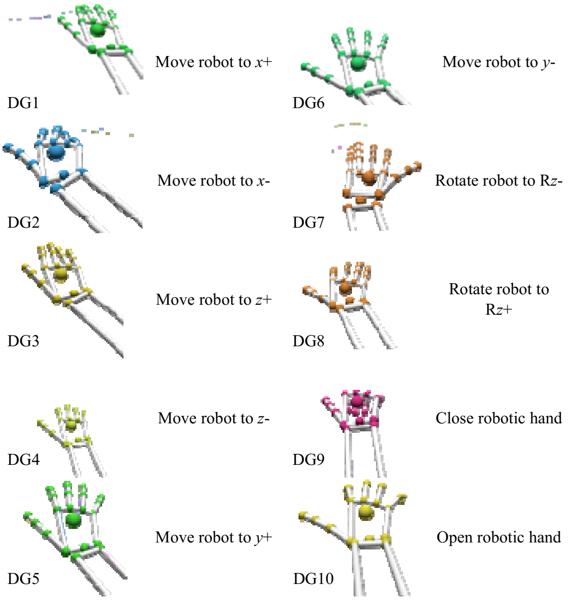

**Fig. 4** Dynamic gestures

## 5 Results and Discussion

The results of the experiment are presented in Tables 6, 7, 8 and 9. In general the SVM architecture provides better RR than the other architectures (ANN and HMM). For the library of static gestures the ANN architecture provided a RR of 96 % while the SVM architecture provided a RR of 97 %. These results are clearly better than the study carried out by Badi et al. [19] who achieved a RR of 86% for a library of just six gestures. The full recognition was not achieved due to inability of the system in differentiate dissimilarities of the gestures. This imposes to the user a more accurate manner in performing the gesture. Another reason for wrong recognition is because the fact that the user performs some involuntary movements which provoke some dissimilarity in gestures. Finally, the occurrence of occlusion between the user hands or among his fingers leads the LMC to make wrong judgments about the scene and outputting wrong features. This effect is clearly visible in the recognition of the SG2 and SG3 which are wrongly recognized between them. Other problem that frequently happens is the recognition of a gesture in the transition between two different gestures. The SG7 is frequently recognized between SG11 and SG12 and vice versa. In fact, the SG7 is recognized because it is really performed during a fraction of milliseconds. In order to cope with this situation, a condition was introduced. This condition consists of validation a static gesture just after it to be continuously recognized during a period of time. In resume, the SVM architecture are able to better differentiate the 12 static gestures proposed in this test case.

In relation to the library of dynamic gestures a RR rate of 94 % was achieved with the HMM architecture while the SVM architecture provided a RR of 96 %. These are good results comparing to the study carried out by Bertsch and Hafner [15], which achieved a RR of 84 % for a library of just seven gestures, and Kurakin et al. [16] that achieved a RR of just 76 %. The hand shake introduce some wrong judgements about a DG this effect is higher with increasing distance between human hand and the center of the LMC. Some gestures trend to be recognized even when there is no intention in perform them, an example of this kind of gesture is DG4. This occurrence is because there is a general tendency to perform a hand moving downward (DG4) when perform any gestures.

It is clear that the SVM architectures provide better outcomes for the gesture recognition system. However, the adopted SVM architecture takes longer processing time than any of the other two architectures leading to slower responses. If the gesture recognition system was not limited by aquisition of data frame, robust static gesture recognitions could be obtained for the SVM architecture at a rate of 10 Hz while the ANN architecture could run at a rate of 80 Hz. Although the SVM architecture provides a slower system, it is effective and quick enough to be used as robot interaction technology.

**Table 6** Confusion matrix for SG – ANN architecture

| | | Recognized gestores | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 | #11 | #12 |
| Performed gestures | #1 | 90 | 5 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #2 | 0 | 80 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #3 | 0 | 5 | 95 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #4 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #5 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #6 | 0 | 0 | 0 | 0 | 0 | 95 | 5 | 0 | 0 | 0 | 0 | 0 |
| | #7 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| | #8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 |
| | #9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 |
| | #10 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 95 | 0 | 0 |
| | #11 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 98 | 0 |
| | #12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |

**Table 7** Confusion matrix for DG – HMM architecture

| | | Recognized gestures | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 |
| Performed gestures | #1 | 83 | 0 | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #2 | 0 | 93 | 0 | 5 | 2 | 0 | 0 | 0 | 0 | 0 |
| | #3 | 0 | 0 | 95 | 5 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #4 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #5 | 0 | 0 | 0 | 2 | 98 | 0 | 0 | 0 | 0 | 0 |
| | #6 | 0 | 0 | 0 | 9 | 2 | 89 | 0 | 0 | 0 | 0 |
| | #7 | 0 | 0 | 1 | 1 | 0 | 0 | 98 | 0 | 0 | 0 |
| | #8 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 96 | 0 | 0 |
| | #9 | 0 | 0 | 0 | 5 | 3 | 1 | 0 | 0 | 90 | 1 |
| | #10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |

**Table 8** Confusion matrix for SG – SVM architecture

| | | Recognized gestores | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 | #11 | #12 |
| Performed gestures | #1 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #2 | 0 | 93 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #3 | 0 | 7 | 93 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #4 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #5 | 6 | 0 | 0 | 0 | 94 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #6 | 7 | 0 | 0 | 0 | 0 | 93 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #7 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| | #8 | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 93 | 0 | 0 | 0 | 0 |
| | #9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 |
| | #10 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 93 | 0 | 0 |
| | #11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 |
| | #12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |

In order to obtain a feasible system with high recognition rates, the same static gesture has to be recognized by one of the recognition architectures for ten times consecutively before the gesture is considered as validated. On the other hand, a dynamic gesture is just validated after it is recognized by the system twice. Thirteen data frames, photos of the gesture, are required to proceed dynamic gesture recognition.

The test bed performed with the presented system to command and program an industrial robot in the execution of an assembly task was carried out successfully. The introduction of conditioning in the recognition of the gestures, i.e. a SG and a DG is just considered as recognized after the system have already identify it five and two times consecutively, respectively, leads to achieve a RR of 100 %. Fig. 5 illustrated the execution of the robotic assembly tasks being the industrial robot commanded by the proposed gesture recognition system.

Instructing the robot with gestures results in a dramatic decrease in programming time, compared to traditional robot programming methods (using a teach pedant). Recurring to the proposed gesture recognition system to program the industrial robot in the execution of the proposed pin assembly task allowed the user to save about 20 % of the time that would be required if the traditional programing system was used. In addition, this system is intuitive and reduced setup time is required.

To the best of our knowledge, approaches similar to ours have never been successfully deployed in real industrial or other scenarios. Thus, we see the main value of our system, as witnessed by our experiment, to operate robustly and solve real industrial problems.

**Table 9** Confusion matrix for DG – SVM architecture

| | | Recognized gestures | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 |
| Performed gestures | #1 | 91 | 0 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #2 | 0 | 90 | 0 | 7 | 3 | 0 | 0 | 0 | 0 | 0 |
| | #3 | 0 | 0 | 93 | 7 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #4 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| | #5 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| | #6 | 0 | 0 | 0 | 10 | 0 | 90 | 0 | 0 | 0 | 0 |
| | #7 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 |
| | #8 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 98 | 0 | 0 |
| | #9 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 96 | 0 |
| | #10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |

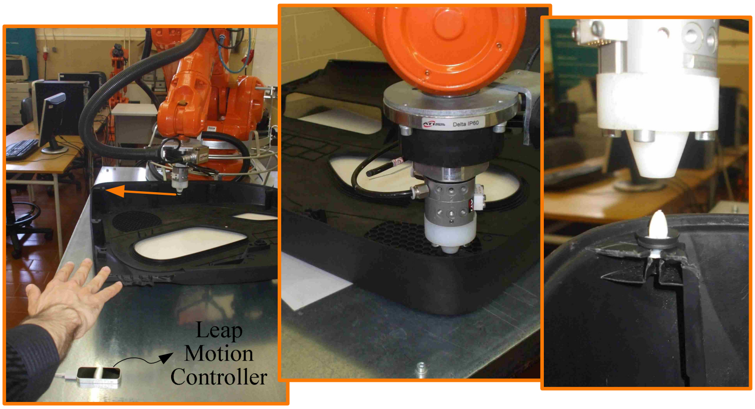

**Fig. 5** Robot executing pin assembly task

# 6 Conclusions and Future Work

A gesture recognition system was proposed to command and program an industrial robot. Static and dynamic gesture features are provided to the system that has implemented three different recognition architectures, i.e. an ANN to static gestures, a HMM to dynamic gestures and a SVM architecture to static and dynamic gestures. All of the architectures provided high RR being the SVM the best one with 97 % and 96 % for a library of 12 static gestures and 10 dynamic gestures respectively. The similarity of the gestures and occlusion problems are preventing the RR improvement.

Finally, as any other system intended for industrial use, an industrial robot equipped with this gesture recognition system has been deployed and tested in an industrial pin assembly task, showing its robustness and effectiveness. We believe that our gesture recognition system constitutes a significant step towards achieving more friendly robots, and that such an approach can ultimately increase competitiveness of manufacturing companies.

**Acknowledgements** This research was supported by FEDER funds through the program COMPETE (Programa Operacional Factores de Competitividade), under the project NORTE-07-0124-FEDER-000060 (COOPERATION).

# References

1. Ren, Z., Yuan, J., Meng, J., Zhang, Z.: Robust Part-Based Hand Gesture Recognition Using Kinect Sensor. IEEE Trans. Multimed. **15**, 1110–1120 (2013)
2. Huang, P.-C., Jeng, S.-K.: Human body pose recognition from a single-view depth camera. In: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2144–2149. IEEE (2012)
3. Seal, A., Bhattacharjee, D., Nasipuri, M., Basu, D.K.: Thermal human face recognition based on GappyPCA. In: 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013), pp. 597–600. IEEE (2013)
4. Kirishima, T., Sato, K., Chihara, K.: Real-time gesture recognition by learning and selective control of visual interest points. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 351–364 (2005)
5. Lambrecht, J., Kruger, J.: Spatial programming for industrial robots based on gestures and Augmented Reality. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 466–472. IEEE (2012)
6. Oz, C., Leu, M.C.: Linguistic properties based on American Sign Language isolated word recognition with artificial neural networks using a sensory glove and motion tracker. Neurocomputing **70**, 2891–2901 (2007)
7. Neto, P., Pires, J.N., Moreira, A.P.: High-level programming and control for industrial robotics: using a hand-held accelerometer-based input device for gesture and posture recognition. Ind. Robot. An. Int. J. **37**, 137–147 (2010)
8. Neto, P., Pires, J.N., Moreira, A.P.: Accelerometer-based control of an industrial robotic arm. In: RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication, pp. 1192–1197. IEEE (2009)

9. Mitra, S., Acharya, T.: Gesture Recognition: A Survey. IEEE Trans. Syst. Man Cybern. Part C Applications Rev. **37**, 311–324 (2007)
10. Yang, J., Bang, W., Choi, E., Cho, S., Oh, J., Cho, J., Kim, S., Ki, E., Kim, D.: A 3D hand-drawn gesture input device using fuzzy ARTMAP-based recognizer. J. Syst. Cybern. Informatics **4**, 1–7 (2006)
11. Yamashita, Y., Tani, J.: Emergence of functional hierarchy in a multiple timescale neural network model: a humanoid robot experiment. PLoS Comput. Biol. **4** (2008)
12. Peng, B., Qian, G.: Online gesture spotting from visual hull data. IEEE Trans. Pattern Anal. Mach. Intell. **33**, 1175–1188 (2011)
13. Badi, H.S., Hussein, S.: Hand posture and gesture recognition technology. Neural Comput. Appl. **25**, 871–878 (2014)
14. Wang, X., Xia, M., Cai, H., Gao, Y., Cattani, C.: Hidden-Markov-Models-Based Dynamic Hand Gesture Recognition. Math. Probl. Eng. (2012)
15. Bertsch, F.A., Hafner, V. V.: Real-time dynamic visual gesture recognition in human-robot interaction. In: 9th IEEE-RAS International Conference on Humanoid Robots, pp. 447–453. IEEE (2009)
16. Kurakin, A., Zhang, Z., Liu, Z.: A real time system for dynamic hand gesture recognition with a depth sensor. In: 20th European Signal Processing Conference (EUSIPCO 2012), pp. 1975–1979 (2012)
17. Zhang, Y., Zhang, L., Hossain, M.A.: Adaptive 3D facial action intensity estimation and emotion recognition. Expert Syst. Appl. **42**, 1446–1464 (2015)
18. El-Baz, A.H., Tolba, A.S.: An efficient algorithm for 3D hand gesture recognition using combined neural classifiers. Neural Comput. Appl. **22**, 1477–1484 (2012)
19. Badi, H., Hussein, S.H., Kareem, S.A.: Feature extraction and ML techniques for static gesture recognition. Neural Comput. Appl. **25**, 733–741 (2014)