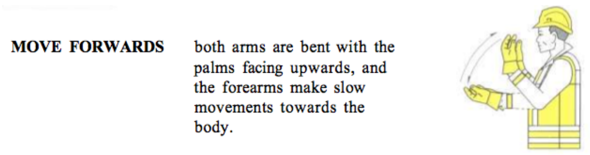
A Proposed Gesture Set for the Control of Industrial Collaborative Robots

P. Barattini[1], C. Morand[2] and N.M. Robertson[2]

Abstract— **Human-Robot Interaction is one of the key challenges in collaborative autonomous robotics. However, no standardised framework allowing either efficient portability to an actual industrial use nor comparison benchmarking exists. This work proposes, implements and evaluates such a set of common ground rules. We present the design constraints between different groups of requirements and a technical solution for automatic recognition using imaging hardware. The Human to Robot and Robot to Human Communications concepts are illustrated on the real industrial scenario: we focus on the definition of a set of gestures for Human to Robot communication in automative manufacturing. The case study outlines the need for a defined set of gestures for establishing a basic communication with the collaborative robot. First, the gestures are designed to respect the social acceptance principle. Second, a gesture recognition algorithm based on Dynamic Time Warping is used to demonstrate the feasibility of discriminating those gestures by automatic processing. Evaluation of our technique shows low confusion and high accuracy with this method.**

I. Introduction

In the Automotive industry despite a high degree of automation in the manufacturing process, the mounting of car parts requires human intervention. Operations such as mounting a trailer hook are performed directly by humans. This kind of lifting of heavy parts by humans is regulated by the Work and Health and Safety regulations because the workers are exposed to the risk of musculo-skeletal injuries. The introduction of a robot assistant can relieve part of the human physical effort and related stress by improving the ergonomics of the mounting process. The adoption of a collaborative robot in the work environment involves cognitive workload and psycho-social factors that must be evaluated. This psychological aspect is of huge importance when making practical decisions on the design of the Human-Robot Interaction (HRI) framework: a tradeoff has to be made between the choice of a technical implementation and what is socially acceptable by the worker. The main contribution of this paper is to determine what this compromise could look like and to evaluate the technical feasibility of the resulting communication gesture set. We show that, in an industrial context, gesture communication is to be preferred; the gestures being defined with respect to both the worker and the robot needs. The gestures are implemented in a real-time computer vision system and evaluated for accuracy.



MOVE FORWARDS both arms are bent with the palms facing upwards, and the forearms make slow movements towards the body.

Fig. 1: An illustrative gesture from the current EU guidelines: a “move forward” command [7]

II. Work environment and HRI requirements

The HRI occurs in a specific environment linked to the main task the autonomous collaborative robot is designed to achieve. This environment dictates the practical and sociological constraints that are presented in this section. In this work we consider the concrete case of a car factory assembly line used in the EC FP7 LOCOBOT project [1]. However, this does not affect the generalization of our results to the more global case of industrial applications. The modern car factory is an evolving environment in which the production line and work areas position are quite often updated and changed according to new production strategies, innovation and also in need of testing new ways. On the one hand the factory physical environment can be very neat and clean, with good illumination, absence of dangerous chemicals, but on the other hand the different work stations are quite cluttered by objects, trailers, pallets, karts, humans, architectural elements such as columns, car bodies moved automatically by the production hovering or slowly descending down to floor level. On top of this, the sound landscape is multifarious and loud, with signals that range from bells to melodies to the honk of electric vehicles and human voices, much more than what the sound experts call the ”cocktail party effect”. The industrial partner requirements exclude any additional device to be worn by the human worker: no microphones wearable for longer distance voice control; no markers or wearable passive tags on the clothes of the worker; no playstation controller, no location sensors armbands or instrumented gloves. All of these are technical options that have been shown to be effective [2], [3] in HRI. This requirement of the industrial partner is justified by the work conditions, the shifting of workers between work areas and tasks and the costs of additional devices, as well by the possible need of commanding the robot by personnel that is not specifically in charge of it and therefore not wearing any specific device.

A. Human-to-Robot Communication

The measurement done on the sound landscape of our test factory tells us that voice control of the robot will be

This work is supported by the EU Commission Grant 260101 LOCOBOT (www.locobot.eu)

[1] P. Barattini is with Ridgeback s.a.s, Turin, Italy (paolo.barattini@sharika.eu)

[2] C. Morand (c.morand@hw.ac.uk) and N.M. Robertson (n.m.robertson@hw.ac.uk) are with Heriot-Watt University, Edinburgh, UK

possible only within a 3m range. It must be pointed out that a 3m range for voice interaction between robot and human is not a strong limitation because it is a tenet that humans have a comfort zone to interact vocally that corresponds to this range. However, the work areas, in which the car parts mounting tasks are to be performed by the mobile robotic assistant, are four times longer.

This means that the gesture communication from the Human to the Robot (HtR) is one valuable option, leaving to the worker the freedom of being further from the robot than the 3m voice control range, while maintaining the capacity of using the robotic assistant through the gesture-visual communication channel. In this case, another requirement is the use of single limb gesture (arm and/or hand and/or fingers) because the human worker could have tools or other devices in her hand, as well as per social acceptance issues and work space occupancy. The use of the right or the left hand is not to be imposed. Despite the on-going research dedicated to Human-Machine Interfaces, Human Computer Interfaces, Human Robot Interfaces and Interaction, no basic set of gestures for industrial interacting robot has been devised and standardized. It must be considered that in the next to come industrial and assistance services scenario there will be many different robots with different roles and possibly produced by different companies or designers working at the same facility. In such a situation, it is very relevant to minimize the cognitive and perceptual workload as well as the training needs for the worker. This can be addressed by defining a set of gestures both easy to learn and re-usable by different robots, these are defined in Section IV.

B. Robot-to-Human Communication (RtH)

The counterpart of a command or signal intended to elicit a behaviour in a cooperative robot is the feedback from it to the human to acknowledge the command. Also a tenet of ergonomics and of Human Machine and Human Robot Interaction is that the system status shall be known to the human operator. In the industrial environment, the feedback of the machines is standardized in ISO Norms related to visual and audio signals: Visual Danger Signals (ISO 11428 et 11429) and Ergonomics (ISO 7331 2005, Danger signals for public and work areas, Auditory danger signals). Another relevant reference is the Directive 2006/42/EC of the european parliament and of the council of the 17th May 2006 on Machinery and amending Directive 95/16/EC.

Though, all of these are not sufficient for the whole HRI need, for which no detailed standard exist.

III. HRI DESIGN GENERALITIES AND CONCEPTS

Since there is no standard for HRI design in a system such as a collaborative autonomous robot, we propose to adopt the following common ground rules. First, we adopt as a stand point the seven principles of Goodrich [4]. Second, we use the following two grids to better characterise our system in relation to the possible roles of the Human in HRI. The first grid lists the possible participants [5]: Supervisor; Operator; Mechanic; Peer and Bystander.

The second regards the modes of cooperation [6]:

1) Robot offers no assistance; human does it all.
2) Robot offers a complete set of action alternatives.
3) Robot narrows the selection down to a few choices.
4) Robot suggests a single action.
5) Robot executes that action if human approves.
6) Robot allows the human limited time to veto before automatic execution.
7) Robot executes automatically then necessarily informs the human.
8) Robot informs human after automatic execution only if human asks.
9) Robot informs human after automatic execution only if it decides to.
10) Robot decides everything and acts autonomously, ignoring the human.

The Human (first grid) in the factory scenario is a Peer and an Operator. Also in relation to the second grid, the robot acts autonomously but in some task (scenario specific) in which there is physical interaction the robot suggests the action (it asks for human cooperation). Regarding the Human side and ergonomics, the following issues must considered and their analysis should contribute to the evolution of the HRI design: Physical and cognitive workload; Emotional components; Comfort; Ergonomics of visual lights signals; Sensory and physical condition of the worker (age, disabilities); Arm reach, visual field, vocal strength.

As a result of the user requirements and the development of an HRI suitable concept we in summary propose that the robot operates via:

1) Vocal commands for short distance interaction (within 3m)
2) Gesture commands are understandable at any distance
3) The gesture and vocal commands are coincident
4) Robot Visual feedback based on 3 LEDs (Red, Yellow, Green) and codification of meaning through frequency and rhythmic patterns.
5) Each gesture (and its coincident vocal command) elicits from the Robot a Visual feedback signal (a pattern).
6) The robot provides audio non-speech feedback e.g. a warning sound. The Robot Status is communicated through LED light patterns and non-speech sounds including special signals for requests of human attention
7) The set of human commands is compact but effective.

IV. BASIC GESTURES SET DESIGN

The European Council Directive 92/58/EEC on the minimum requirements for the provision of safety signs at work includes some hand signals. This directive is implemented through the National Regulations [7]. This set of signals is intended for the communication between two workers, one of them controlling a machine, such as a crane or a forklift, the other providing directions.

Many of these signals (e.g. see Figure 1) are to be performed with two hands, they are in some way dramatic, and they request an envelope quite wide (space around the person

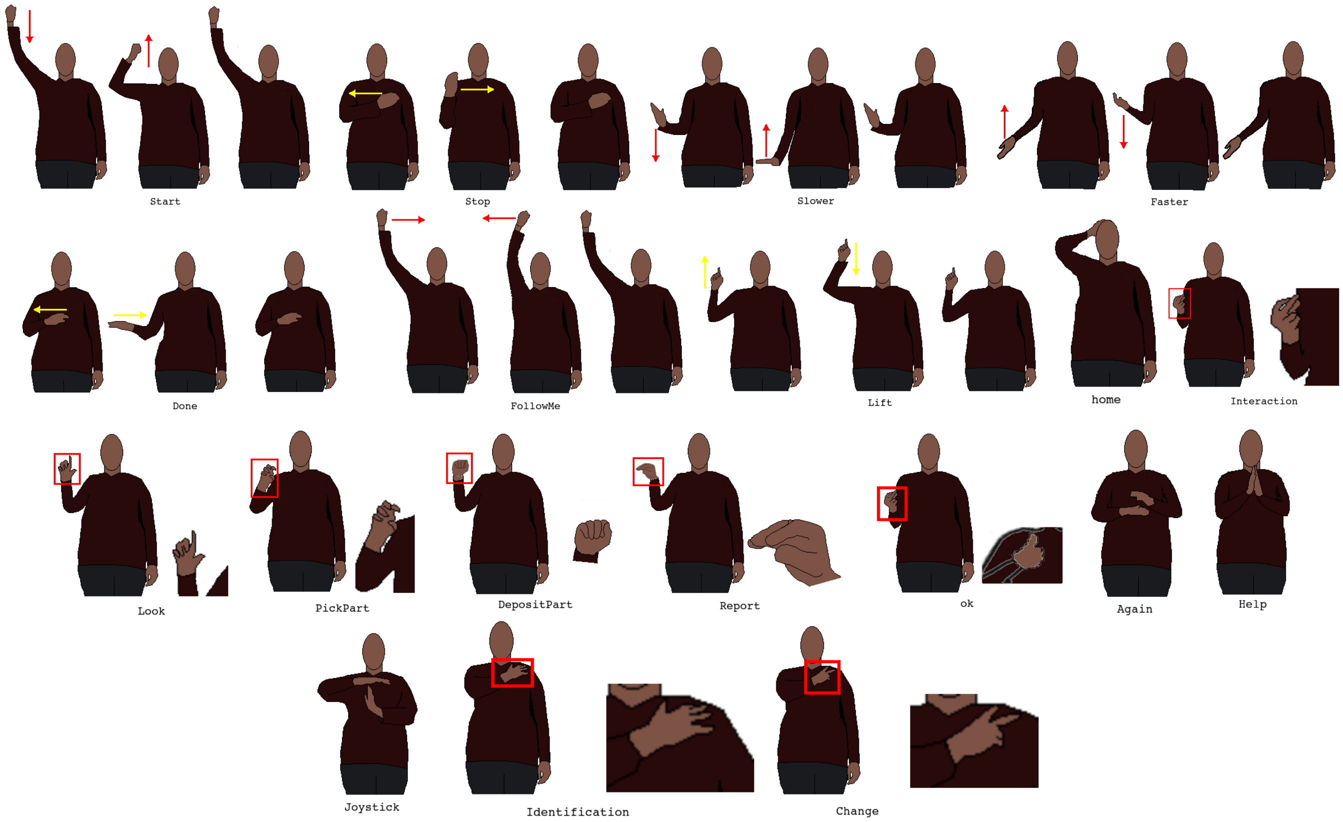


Fig. 2: Our proposed gesture set based on the established HRI principles described in the text.

that is free of obstacles). So, they do not suit our HRI requirements (see Section II). Nevertheless, they point to the fact that in the future the regulation will have to take into consideration human-robot gesture communication.

There is much prior work in the literature related to gesture recognition ([19], [20]), but we note there is no standardised set of gestures is available. Furthermore, most of the lab developed gestures were finalized to experiment the capacity of the technology, and not to perform tasks in a real world industrial environment with safety and environmental constraints. For example, a recent paper by Burger [8] briefly reviews relevant work in this area.

A. *Design requirements for a gesture*

The gestures must be easy to make, as close as possible to the common use of finger/hand/arm gestures(i.e. “natural”), clearly distinguishable one from another, easy to remember, minimizing training time, different from movements done to perform work tasks, different from gesticulation done while talking, socially acceptable, minimizing the cognitive workload.

There are many classifications of Gestures (e.g [9], [10], [11], [12], [13], [14]) that are useful in Human Robot Interaction. Each of the gesture categories has a different relation with the speech channel of communication, with regards to its capacity of expressing meaning independently from it. So gestures can be ordered according to their degree of independent understandability without speech. In Kendons Continuum [10] the types of gesture are listed in increasing order of independency from speech:
Gesticulation > Language-Like > Pantomimes > Emblems > Sign Language.
Hence a sign language such as the American Sign Language is rich in syntactic and semantic features at a degree that does not need the contemporarily use of spoken language and would seem to be the best type for gesture-only communication. However, such a sign language is quite complicated and requires a great learning workload, so it is not suitable for use to express simple commands for robots in industrial environment.

Another way of categorizing gestures is proposed by McNeil [9]:

- Iconic: represents images of concrete/abstract entities and/or actions, with resemblance to the event or objects, i.e. they have semantic connection to it.
- Metaphoric: are related more to the representation of abstract entities or concepts rather than concrete objects
- Deictic: the gestures of the kind “pointing index”. Can be done also with other body parts. Typically refers to entities/actions/objects present in the environment of the person acting, but can also be abstract.
- Beats: so called because the movement of the limb/hand is rhythmic like if it were beating time. Typically is what happens in Gesticulation, with no semantic correspondency to the speech.

Based on the requirements highlighted in sections II and III, and considering that the gesture could be done also in absence of speech (3 meters range limitation, II), we have to design gestures that are mainly Iconic or Deictic, or in other words Pantomimic/Emblems.

B. Gestures set

The next step in the design of the gesture set is to consider how the gestures are connected to the action to be performed by the robot that is nested in the work processes; this creates a branching network for the commands logic. Furthermore, to minimize the worker learning phase, the number of gestures to be used has to be relatively small. Based on the real case of the LOCOBOT project, we were able to define the essential commands to give to a collaborative autonomous robot. It resumes to a basic set of 12 gestures. Another supplementary set of 8 gestures is added to extend the capacity of the robot in other tasks and work areas.

Those gestures, along with their signification for the robot, are presented in Figure 2.

Two of these gestures have a different role than the others: "identification" and "change". Upon turning on the robot with a physical button, it will not answer to any command. This is because of the responsibility and safety requirements and constraints that tell us that only one person has to be in charge of the robot, and this is not possible without prior identification. This means that only the "identification" gesture command, or vocal or fused Gesture/vocal command will work. Only after this procedural step, the other commands will be available. Conversely, the robot must be technically capable of performing an identification of the responsible worker and tracking her continuously so to be able to answer and act on the basis of her only gestures. Sometimes, for example at the end of the work-shift or production process steps, it is needed to change the worker in charge of the robot. This entails immediately the need of another command: "change" identification.

Dynamic gestures were the preferred choice because they are more natural: they have to command actions and movements and so dynamic means semantically related to the objective. They can communicate additional meaning. Actually after selecting one arm gestures for about seven commands ("start", "stop", "slower", "faster", "done", "follow me", "done"), the repertoire of simple natural dynamic gestures of the arm came to an end. So to implement the other commands we resumed to combined movement of arm, hand and fingers.

Last but not least at all Industry users claim that for a robot to be valuable for the use an Industrial Production Process should have a failure rate that is null. The technical choices have to be benchmarked on the basis of all of this environmental and requirements background, even if a failure rate of zero is in practice unachievable.

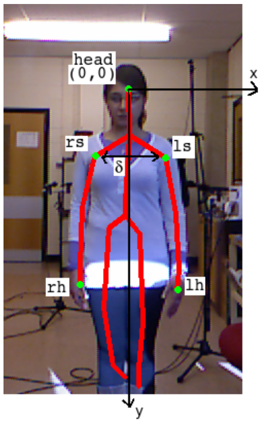


Fig. 4: Coordinate system used to normalize the feature vector. The skeletal points are automatically extracted in real-time from the images.

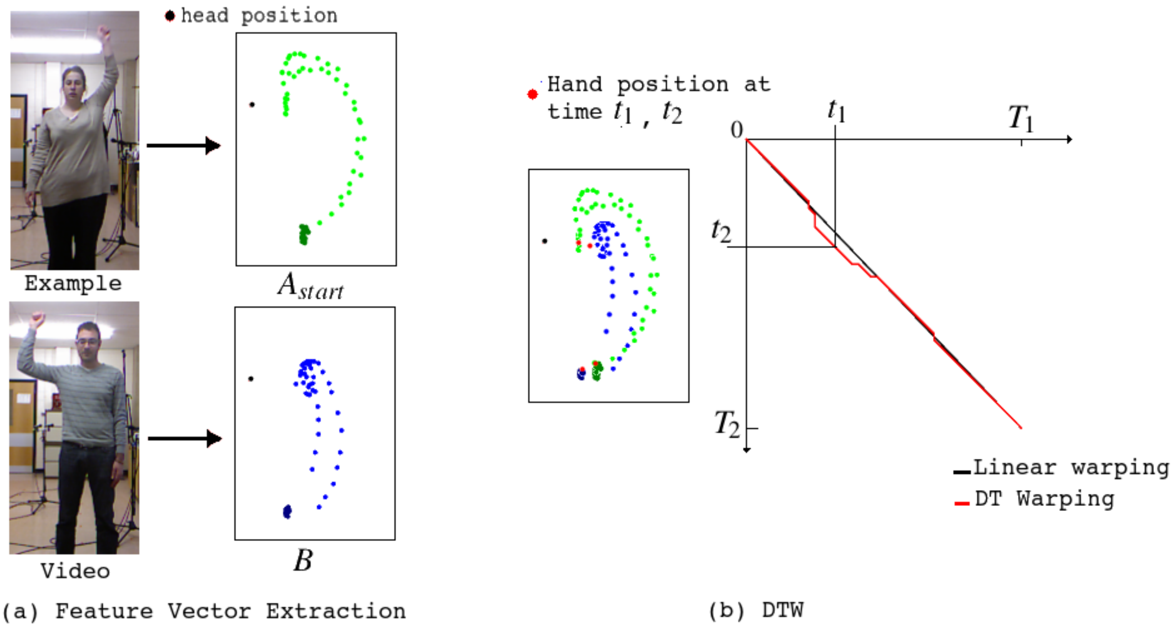
V. Gesture recognition

Having define the set, the next step is to make a first evaluation on the capacity of a computer to distinguish between the gestures. Our aim is to provide a simple tool to analyze the proposed gestures (and possibly others) rather than solve the recognition problem for this particular set. Hence, when making choice in the design of the algorithm, we try to choose the simplest alternative.

A. Gesture recognition algorithm

The set presented in Figure 2 contains different types of gestures, namely single and double handed dynamical gestures and single and double handed pose gestures, which can furthermore be combined with hand pattern gestures. No general algorithm can cope with all this variety. However, it can be noted that all of these gestures can primarily be described by a specific position of the hands with respect to the position of the head. This carachteristic can be used for a first rough recognition of algorithm. The principle of the proposed algorithm is illustrated in Figure 3 and described below.

Automatic hand/head segmentation and tracking is a commonly studied problem in the literature, approaches ranging from the color-based "camshift" algorithm [15] to the estimation of the skeleton positions like in [16]. In this paper, we use the skeleton extraction approach as provided by the Microsoft Kinect SDK (see V-B, dataset acquisition). The feature vector A describing a gesture is the time series of the normalized positions (x,y) of the hands as described in equation (1) (see Figure 3(a)). The axis system used is described in figure 4.



(a) Feature Vector Extraction

(b) DTW

Fig. 3: Gesture Recognition Algorithm principle illustrated on the start gesture. Hands positions are extracted for both the example and the test gesture sequence (a), then a point-by-point distance is computed, the point association being determined by the dynamic time warping function (b).

$$A = (a_t)_{t\in[0,T]}\ ,\ a_i = (lh(a_i), rh(a_i))$$

$$lh(a_i) = \begin{pmatrix} -(x_{lh}-x_{head})/\delta \\ (y_{lh}-y_{head})/\delta \end{pmatrix}_i \ ,\ rh(a_i) = \begin{pmatrix} (x_{rh}-x_{head})/\delta \\ (y_{rh}-y_{head})/\delta \end{pmatrix}_i \qquad (1)$$

$$\delta = d(ls,rs) = \sqrt{(x_{ls}-x_{rs})^2+(y_{ls}-y_{rs})^2}$$

where lh stands for left hand, rh: right hand, ls: left shoulder, rs: right shoulder. $t = 0$ (resp. $t = T$) is the beginning (resp. end) of the gesture. δ is the euclidian distance between the two shoulders used as normalizing distance to take into account the inter-people variations in size.
Note that the position of the left hand is mirror w.r.t. the y-axis. This is to sustain the requirement that each one-handed gesture can be performed indifferently with the right or the left hand.

As dynamic gestures can be done at different paces by different workers, we choose to use a simple version of the Dynamic Time Warping algorithm to determine the distance between two feature vectors [17], [18]. The recurrence steps used in practice to determine the warping function (Figure 3(b)) and further determine the distance between $A = (a_{t_1})_{t_1\in[0,T_1]}$ and $B = (b_{t_2})_{t_2\in[0,T_2]}$ are given by the following equations.

- Initial Condition : $g(t_1 = 1, t_2 = 1) = d_m(a_1, b_1)$
- Dynamic Programming Equation:

$$g(t_1,t_2) = d_m(a_{t_1}, b_{t_2}) + min \left\{ \begin{matrix} g(t_1, t_2-1) \\ g(t_1-1, t_2-1) \\ g(t_1-1, t_2) \end{matrix} \right\} \qquad (2)$$

- Distance: $D(A,B) = \frac{1}{T_1+T_2} g(T_1,T_2)$

The distance d_m between two elements of A and B is define so that no a priori is done on the leading hand used to make the gesture. It is given by:

$$d_m(a_i,b_i) = min \left\{ \begin{matrix} d(lh(a_i),lh(b_i)) + d(rh(a_i),rh(b_i)) \\ d(lh(a_i),rh(b_i)) + d(rh(a_i),lh(b_i)) \end{matrix} \right\}$$

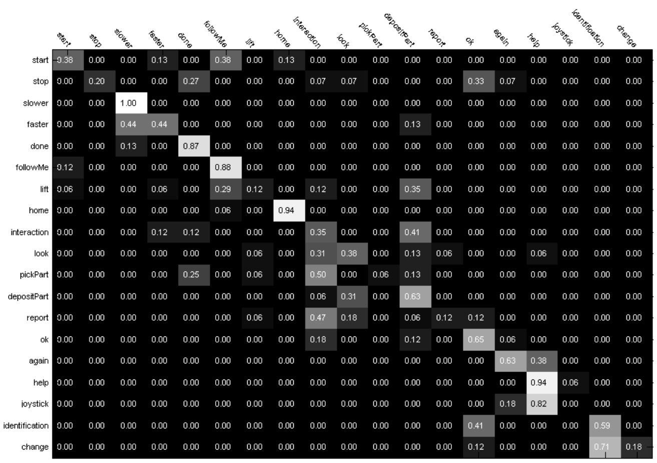
For performing the recognition algorithm, we make the hypothesis that the proposed gestures in Figure 2 are performed completely. In particular, it means that gestures begin and end with a passage to the resting position - both arms lying along the body. Gesture segmentation can then be performed thanks to this hypothesis.
Let the example set be $\{A_{start}, A_{stop}, ..., A_{change}\}$ where A_g is the feature vector example representing gesture g. A performed gestured B is recognised as gesture G when $D(B, A_G) < T$ with T a experimentally defined threshold and

$$G = \underset{g\in\{\text{start, stop, ..., change}\}}{argmin} (D(B,A_g)) \qquad (3)$$

In the following, the example set is extracted from only one recording of each gesture rather than learnt from a set of recordings.

B. Evaluation of automatic gesture recognition

The 19 different gestures were performed 3 times by 6 different persons and stored individually. To record, we choose to use a ©kinect device. This is indeed the kind of low-cost ready-to-use solution that one can realistically expect to find on a robot. Note that even if this device is sufficient to proove the feasibility of the gesture recognition, it has a limitation of 4-meters range. However, the approach



	start	stop	slower	faster	done	followMe	lift	home	interaction	look	pickPart	depositPart	report	ok	again	help	joystick	identification	change
start	0.38	0.00	0.00	0.13	0.00	0.38	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
stop	0.00	0.20	0.00	0.00	0.27	0.00	0.00	0.00	0.07	0.07	0.00	0.00	0.00	0.33	0.07	0.00	0.00	0.00	0.00
slower	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
faster	0.00	0.00	0.44	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00
done	0.00	0.00	0.13	0.00	0.87	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
followMe	0.12	0.00	0.00	0.00	0.00	0.88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
lift	0.06	0.00	0.00	0.06	0.00	0.29	0.12	0.00	0.12	0.00	0.00	0.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00
home	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
interaction	0.00	0.00	0.00	0.12	0.12	0.00	0.00	0.00	0.35	0.00	0.00	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00
look	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.31	0.38	0.00	0.13	0.06	0.00	0.00	0.06	0.00	0.00	0.00
pickPart	0.00	0.00	0.00	0.00	0.25	0.00	0.06	0.00	0.50	0.00	0.06	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00
depositPart	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.31	0.00	0.63	0.00	0.00	0.00	0.00	0.00	0.00	0.00
report	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.47	0.18	0.00	0.06	0.12	0.12	0.00	0.00	0.00	0.00	0.00
ok	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00	0.00	0.12	0.00	0.65	0.06	0.00	0.00	0.00	0.00
again	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.63	0.38	0.00	0.00	0.00
help	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.06	0.00	0.00
joystick	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.82	0.00	0.00	0.00
identification	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.41	0.00	0.00	0.00	0.59	0.00
change	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.71	0.18

Fig. 5: Confusion matrix showing the results of comparisons among the gestures in the proposed set.

is not sensor specific and can be used with more accurate Photonic Mixer Devices (PMD) that delivers a larger range. In the following, the example set is chosen as the set of gestures performed by one individual that is the nearer to the original pattern gesture presented in figure 2. Indeed, in the recording phase, the persons were only taught the general pattern of the gestures but let free to perform them as they feel the most confortable. As a result, there can exist a intra-gesture variety in terms of amplitude and repetition for instance.

Figure 5 presents the confusion matrix obtained in our experiments. The matrix is sparse and quasi-diagonal thus demonstrating the capacity effectively to distinguish between gestures. All the gestures are presented here, separated on the sole criterion of the dynamic position of the hands. However, two gestures from the dataset may only differ by hand pattern. It is the case for the groups {identification, change} and {interaction, look, pickPart, DepositPart, report, ok}. As expected, the confusion inside the group is more elevated. But there is a low confusion with other gestures. So, when dealing with a practical implementation, a two-step algorithm can be designed: first, determining the dynamical position as we have done, second, for the specifically identified group, apply a hand pattern recognition algorithm. Identically, faster and slower have the same dynamic pattern and differ by the positioning of the hand palm up and palm down. In this last case, poor results can be expected from the use of the visual signal. The gestures have to be made more obvious, needing more training from the workers, for instance by imposing that the hand for faster has to be at the level of the upper torso and slower at the lower body.

VI. Conclusion and Future Work

The development of an effective HRI involves research and design of the human gestures to develop a basic set of as "natural" as possible commands through arm and hand movements. The interplay of this aspect with the technical assessment of the available scientific tools can validate possible candidates for standardisation of human robot gesture based interaction, whereas gestures must also be "understandable" by the robot minimizing interpretation error rates. Our future work will focus on (a) incorporating audio commands which are concurrent with the gestures; (b) comparing state-based techniques to the continuous method presented here; (c) expanding the gesture set to include fingers; (d) Testing *in situ* the gesture set with factory workers.

References

[1] http://www.locobot.eu/

[2] S.Mitra, "Gecture recognition: A survey", IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews, vol. 37, N. 3, May 2007, p311-324.

[3] M.Urban, P. Bajcsy, R. Kooper and JC. Lementec, Recognition of Arm Gestures Using Multiple Orientation Sensors: Repeatability Assessment, IEEE Intelligent Transportation Systems Conference Washington, D.C., USA, October 3-6, 2004.

[4] M.A. Goodrich and D.R. Olsen, Jr. Seven Principles of Efficient Interaction. Proceedings of IEEE International Conference on Systems, Man and Cybernetics, October 5-8, 2003, pp 3943-3948.

[5] J. Scholtz, M. Theofanos and B. Antonishek, Theory and evaluation of human robot interactions", in 36th International Conference on Systems Sciences, Hawaii, IEEE, 2002.

[6] M.A. Goodrich and A.C. Schultz. Human-Robot Interaction: A Survey. Fondations and Trends in Human-Computer Interaction, 1(3), 2007, pp 203-275.

[7] The British Health and Safety Executive (Safety Signs and Signals) Regulations second edition 2009 `http://books.hse.gov.uk/hse/public/saleproduct.jsf?catalogueCode=9780717663590` Retrieved 29 March 2012

[8] B. Burger , I. Ferrane, F. Lerasle, and G. Infantes, Two-handed gesture recognition and fusion with speech to command a robot Autonomous Robots Volume 32, Number 2, 2012, pp 129-147.

[9] McNeill, D. (1992). Hand and mind: what gestures reveal about thought. Chicago, USA: University of chicago press.

[10] Kendon, A. (1988). How gestures can become like words. In Potyatos, F. (ed), Crosscultural perspectives in nonverbal communication, p 131-141. Toronto, Canada: Hogrefe.

[11] Cadoz, C. 1994. "Le geste canal de communication homme-machine. La communication instrumentale" Sciences Informatiques, numeero speecial: Interface homme-machine. 13(1): 31-61.

[12] P. Ekman,W.V. Friesen The repertoire of non verbal behavior : categories, origins, and coding. Semeiotica, 1, 49-98, 1969.

[13] Efron, D. (1941). Gesture and environment: A tentative study of some of the spatio-temporal and linguistic aspects of the gestural behavior of eastern Jews and southern Italians in New York City, living under similar as well as different environmental conditions. New York: Kings Crown Press.

[14] Nespoulos, J.L., Roch Lecours, A., (1986). Gestures: nature and function. In: Nespoulos, J.L., Perron, P., Roch Lecours, A., The biological foundations of gestures: motor and semiotic aspects, p 49-62. Hillsdale, New Jersey, USA: Lawrence Erlbaum Associates.

[15] G.R. Bradski. Real-Time Face and Object Tracking as a Component of a Perceptual User Interface", Fourth IEEE Workshop on Applications of Computer Vision (WAC'98), 1998.

[16] Z.L. Husz, A.M. Wallace and P.R. Green. Behavioural Analysis with Movement Cluster Model for Concurrent Actions. EURASIP Journal on Image and Video Processing. 2011.

[17] H. Sakoe and S. Chiba. Dynamic Programming Algorithm Optimization for Spoken Gesture Recognition. IEEE Transactions on Accoustics, speech and signal processing, vol. ASSS-26, no 1. 1978.

[18] A. Corradini. Dynamic time warping for off-line recognition of a small gesture vocabulary. Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 2001. Proceedings. IEEE ICCV Workshop on. pp 82 - 89.

[19] M.B. Holte, T.B. Moeslund, P. Fihl. View-invariant gesture recognition using 3D optical flow and harmonic motion context. Computer Vision and Image Understanding 114 (2010) 13531361.

[20] Van den Bergh, M. Carton, D. ; De Nijs, R. ; Mitsou, N. ; Landsiedel, C. ; Kuehnlenz, K. ; Wollherr, D. ; Van Gool, L. ; Buss, M. Real-time 3D hand gesture interaction with a robot for understanding directions from humans. RO-MAN, 2011 IEEE. p357 - 362.