Rich and Robust Human-Robot Interaction on Gesture Recognition for Assembly Tasks

Gi Hyun Lim, Eurico Pedrosa, Filipe Amaral, Nuno Lau,
Artur Pereira, Paulo Dias, José Luís Azevedo, Bernardo Cunha, and Luis Paulo Reis

***Abstract*— The adoption of robotics technology has the potential to advance quality, efficiency and safety for manufacturing enterprises, in particular small and medium-sized enterprises. This paper presents a human-robot interaction (HRI) system that enables a robot to receive commands, provide information to a human team mate and ask them a favor. In order to build robust HRI system based on gesture recognition, three key issues are addressed: richness, multiple feature fusion and failure verification. The developed system has been tested and validated in a realistic lab with a real mobile manipulator and a human team mate to solve a puzzle game.**

I. Introduction

The adoption of robotics technology has the potential to improve quality, efficiency and safety for manufacturing enterprises [1]. With an increasing number of robots for manufacturing, robots are required to deploy beyond traditional automated tasks in small and medium-sized enterprises (SMEs) where humans and robots work together for production. Their environment is typically less structured and demands higher flexibility than large-scale or mass-production industries [2]. Collaborating in a workspace, all the participants need to interact and communicate to each other to complete tasks [3]. Since many industrial environments are noisy where machines continue to whir, verbal communication is difficult for humans and impractical for robots [4]. Thus, gestures which are visible bodily actions to communicate particular messages have been considered as a means of communication.

The EuRoC research project has launched industry-relevant challenges to exploit synergies among all the actors of the value chain in manufacturing and servicing [5]. Especially the challenge 2, named *Shop Floor Logistics and Manipulation*, addresses *Logistics and Robotic Co-Workers* scenarios of the SRA2009. As an evaluation of *Stage II* a Free-style, a puzzle solving scenario has been designed team to test and validate the developed system in a realistic lab before being applied in real industrial environments by the TIMAIRIS. The lab is configured with a real KUKA LWR manipulator mounted on an omniRob mobile base exposing an interface on Robotic Operating System (ROS) and two tables: a pick up station and an assembly station. The scenario is to assemble pentominoes (P5s, see Fig. 4) which are polygons by connecting five equal-sized cubs face-by-face, to fill a given space with cooperation between a human and a robot. The robot, or the human, fetches a P5 from the tool shelf to the workbench and assembles the P5 into a puzzle space to complete the puzzle. The task has four objectives: perception, manipulation, planning and interaction. To address the requirements and achieve the objectives, a skill-based architecture has been developed to solve the logistics and manipulation tasks [6], [7]. In particular, this paper addresses the human-robot interaction system with gesture recognition.

There are three key issues on robust human-robot interaction based on gesture recognition in industry settings. First, HRI system needs rich representations especially for SMEs. They are commonly low-volume manufacturers or small batch producers whose products continue to rapidly evolve. However, it is difficult and a burden for SMEs to change or add new features such as materials and tools. To solve this problem, interaction graphs are proposed to encode any command fixed number of symbols (see Sec. III-B). Second, there is no single distinctive feature to classify all the hand gestures [8]. Even a feature trained in deep neural networks with high accuracy has confusion between specific gestures. In this paper, multiple features such as two features from deep neural networks and a contour-based feature are integrated (see Sec. IV). Lastly, even with a multiple feature fusion method, there are still many false classification results. To increase robustness, two steps of verification methods are applied such as confidence-based verification and word correction (see Sec. V).

II. Related work

During recent decades, robotics has been integrated to boost manufacturing for the purpose of performing repetitive and dangerous tasks such as assembly tasks. For SMEs whose environments are unstructured and cluttered, a robot becomes an assistant or collaborator that works literally hand-in-hand with its human counterparts. Here, the robot needs to communicate with humans to take an order, to ask confirmation, and to reply to a question or to ask for a human task. For example, the human wants to know the assembly plan, asks to bring parts and/or tools to a workbench from far pickup station, and demands to place them in a position or on a hand.

To adapt to SME environments, gesture based human-robot interaction (HRI) systems have been applied, e.g. [9], [10]. In [9], three types of gestures (beckon, give and shake hands) have been demonstrated to interact between a human and a humanoid robot. Gerard et al. [10] presented a HRI

All authors are with IEETA - Instituto de Engenharia Electrónica e Telemática de Aveiro, Universidade de Aveiro. Email: {lim, efp, f.amaral, nunolau, artur, paulo.dias, jla}@ua.pt, mbc@det.ua.pt, lpreis@dsi.uminho.pt

system that recognizes four gestures: wave, nod, negate and point at. For the wave gesture, the expected response is waving back. The nod and negate gesture are for assent and dissent, respectively. The point at gesture is for designating an object laying between a human and a robot. To the best of our knowledge, gesture recognition systems in HRI have focused on small number of implicit interactions such as pointing and handing over. To ask to fetch objects in a cluttered and unstructured environment, HRI system needs to have a rich vocabulary to specify an object.

For robust gesture recognition, a single feature is not sufficient [11]. [12] reports that the recognition rate of deep neural networks is higher than 95%. However, there are frequent misidentifications between '2' and '3' of one-hand gestures (see Fig. 2), since both are similar in terms of convolutional image filters. On the other hand, a contour-based feature [13] is good to distinguish the two gestures, but not good for some others such as between '1' and 'thumbs up', since both gestures are achieved by a closed fist held with one finger extended. To remove noisy features, multiple feature fusion methods deal with the selection and combination of features. Several strategies such as parallel fusion [14] and feature selection, ranking, and feature combination [8] have been surveyed. In this paper, three features: two from deep neural networks and a contour-based feature are fused together to obtain a robust gesture recognition result.

In a real environment, there is one significant difference with training phase. During the transition between two gestures, some image frames are false or spurious. Those images are discarded in training phase, since humans cannot classify them. However, those are also classified as one of classes in the trained models. To remove False Positives (FPs), several verification methods are proposed [15], [16], [17] Ishan et al. [15] proposed a sequential verification to determine whether a sequence of image frames is in the correct temporal order or not. Dirk et al. [16] applied an object verification step after object detection and pose refinement. Using this approach, they could find wrong registration results. In [17], a geometric verification step has been added for object localization and recognition by checking the spatial consistency. In this work, two verification methods: confidence-based verification and word correction are applied to refine the gesture recognition result and to find a proper command that is in a codebook from misspelled words.

III. System architecture

A. HRI system architecture

Figure 1 shows the detailed architecture of HRI. The developed system is composed of an interaction graph, hand dataset and five modules: *Human detection*, *Frame and feature extraction*, *Gesture recognition*, *Sentence inference* and *Multi-modal expression*. Each module is structured as a ROS node and fully integrated in the platform, and the system relies on *Open Source Computer Vision* (OpenCV)

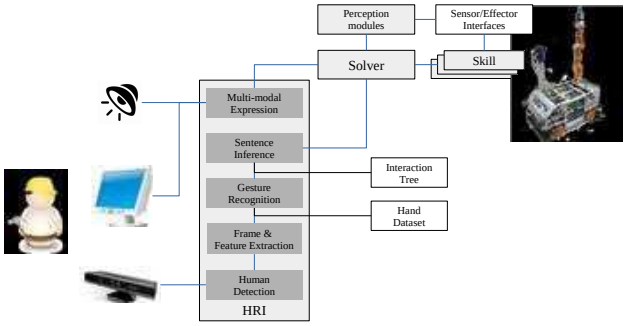


Fig. 1. HRI architecture.

[1] and *Point Cloud Library* (PCL) [18] [2] libraries to handle RGB-D data and their corresponding 2D images [19].

B. Interaction graph

Level	Description
1	Yes/OK
1	No
1	Get me next pentomino
2	Place in the table
2	Place aligned with final position
2	Place in my hand
2	Place in the puzzle
2	Show me a next pentomino
2	Show me available pentominoes
2	Show me final solutions sequencely
2	Show me assembly planning
2	Show me a place to put the pentomino
2	Get me the pentomino : {F, I, L, N, P, T, U, V, W, X, Y, Z }

TABLE I

24 COMMANDS FOR THE PUZZLE ASSEMBLY.

Table I shows all possible commands from a human to a robot for the scenario. 24 commands are defined in advance. Additionally, the robot not only provides information to show what it is doing but also requests two kinds of tasks to demonstrate mixed-initiative interaction between human and robot [20], since robot initiative can make the pace of interaction higher and the reaction to attention faster [21].

To enable its use in noisy environments, such as industry, human commands are given via hand gestures as a sign language [4]. 8 one-hand gestures are selected as basic symbols, as shown in Fig. 2, and 24 commands are composed by concatenating the gestures as sentences.

Identically to a human language, where unlimited sentences can be composed of a small and fixed number of symbols, an interaction graph and pentomino table are designed to construct 24 sentences using 8 gestures. The interaction graph consists of 13 kinds of commands with 8 gestures, as shown in Fig. 3, while 12 P5s are designated by combination of 2 gestures out of 5 gestures, as shown in Table II. For example, the sequence of two gestures: *2* and *1* means P5 T. In Table II, 12 P5s are evenly spread out and the diagonal entries are left blank. That provides robustness on sentence inference from a sequence of gestures, and enables

[1] http://opencv.org/.
[2] http://pointclouds.org/.

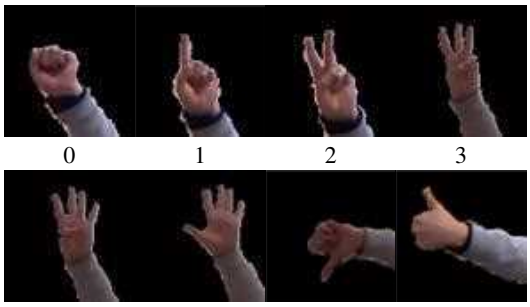


Fig. 2. 8 one hand gestures and their corresponding IDs.

identification of a new gesture, as there are no sequences that repeat the same symbol in succession. Figure 4 shows 12 P5s, which are polyhedrons formed by joining face by face five identical cubes on a single plane. A set of 12 different P5s may be used for a puzzle. For solving the puzzle game, six colored and six uncolored wooden P5s are used. They are identified by an alphabet letter resembling their shape.

	1	2	3	4	5
1		T		L	
2	I		Y		V
3		W		N	
4	P		F		Z
5		X		U	

TABLE II
PENTOMINO TABLE

C. Gesture recognition pipeline

Figure 5 shows a proposed pipeline for gesture recognition. When the system receives a steam of RGB-D data as inputs, the gesture recognition pipeline is launched which contains four modules: *Human Detection*, *Hand Detection*, *Feature Extraction* and *Gesture Recognition*.

IV. Gesture representations

To recognize gestures, two kinds of features are used: convolutional representation from deep learning and a contour-based hand feature. The former is a cutting edge method, which gives human level recognition performance in general [22], [23]. However, there are some misidentifications between *2, 3* of one hand gestures, since convolution representations might extract distinctive features of whole image region, but the image of three gestures is mainly similar and only parts are different. The later is good to detect the number of fingers [13], [24]. However, it requires a well-segmented hand image and its pose. Still, there are some misidentifications between *0, thumbs down* and between *1, thumbs up* in online running. Since they are complementary, merging the two methods provides better performance for all one hand gestures classes in the challenge environment.

A. Convolutional representations

Many outstanding results in computer vision are currently achieved by deep learning methods that attempt to learn feature representations [22], [25], [26]. One of the key success factors in deep learning is its hierarchical structure, as with convolutional deep networks able to learn rich representations. They are grounded in high-precision gradient descent methods for training.

Convolutional neural networks (CNNs) are multi-layered neural networks mainly applied on recognizing visual patterns directly from images. We utilize the well-known CNN of [22]. 8 one-hand gesture models are trained by the fast and standard stochastic gradient descent algorithm. During training, a data layer fetches the images and labels and passes it through multiple layers, in detail: 5 convolution layers, 7 rectified-linear unit (ReLU) layers, 2 local response normalization (LRN) layers, 3 max-pooling layers, 3 fully connected layers, 2 dropout layers and a SoftmaxWithLoss layer for classification.

B. Contour-based hand feature

After detecting the hand shape, we represent it as a contour-based hand feature, as shown in Fig. 6. Such a shape representation, which plots the relative distance between each contour vertex and the hand center point, has been successfully used for hand gesture recognition [13]. First, the system detects the contour of the hand, and then finds the largest circle in the contour. The center point of the circle is used as the hand center point, drawn as rounded rectangles in Fig. 6. The start point and the end point of contour histogram, drawn as solid line in Fig. 6 are defined as the two farthest points of the contour on the image boundary by discarding the first downslope and the last upslope, drawn as large circles and small circles in Fig. 6, respectively. To discard the contour of arm, the first and last slopes are removed from the plot.

To recognize a hand gesture with the contour-based feature histogram matching was used, that is, the histogram of input hand is classified as the class with which it has the minimum Bhattacharyya distance [27]:

$$c = \arg\min_{c} D_{Bhattacharyya}(H, T_c), \qquad (1)$$

$$D_{Bhattacharyya}(H, T_c) = \sqrt{1 - \sum_i \frac{\sqrt{H \cdot T_c}}{\sqrt{\sum_i H \cdot \sum_i T_c}}}, \qquad (2)$$

where H is the input histogram; T_c is the template of the class c. Bhattacharyya distance measures the amount of overlap between two discrete or continuous distributions.

V. Sentence Inference

The HRI system needs to construct a sentence from the continuous gesture recognition results. The two kinds of results from CNNs and contour-based hand gesture recognition are weightily merged according to their average values.

A. Confidence-based verification

To construct a sentence from a sequence of gesture recognition results, a confidence-based method is firstly applied to concatenate a sequence of symbols robustly [28]. An interval-counter (γ) for each gesture is defined on the basis of the *confidence law of inertia*, whereby a recognized symbol is assumed to persist unless there is confidence to believe

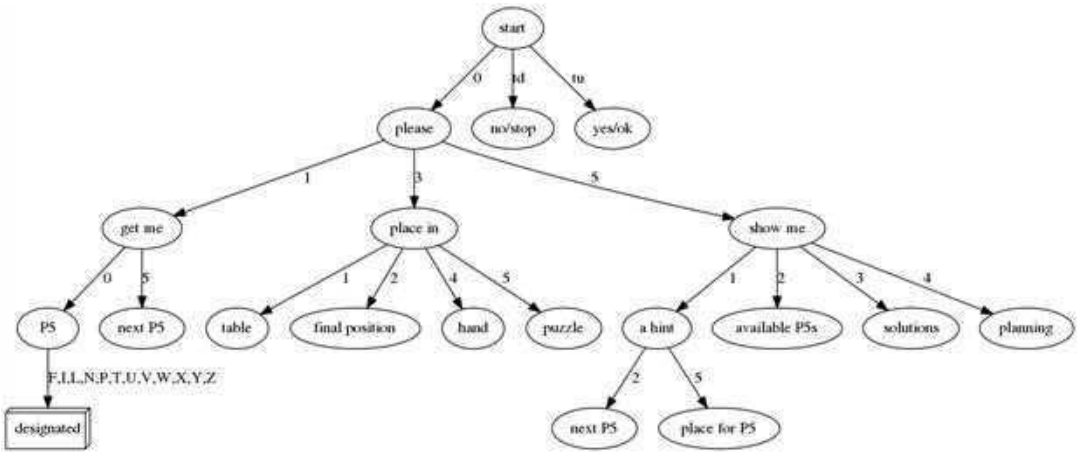


Fig. 3. Interaction graph for 24 commands.

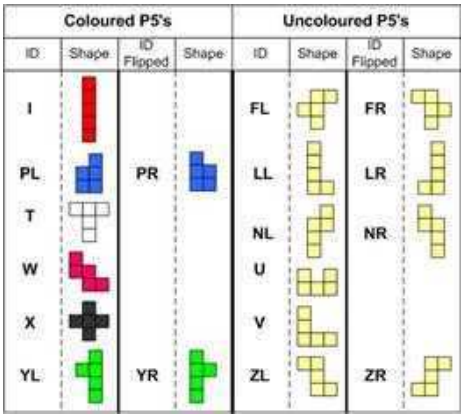


Fig. 4. Shapes and alphabetic IDs of 12 pentominoes.

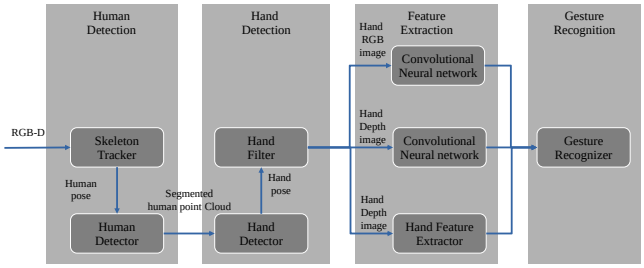


Fig. 5. Gesture recognition pipeline.

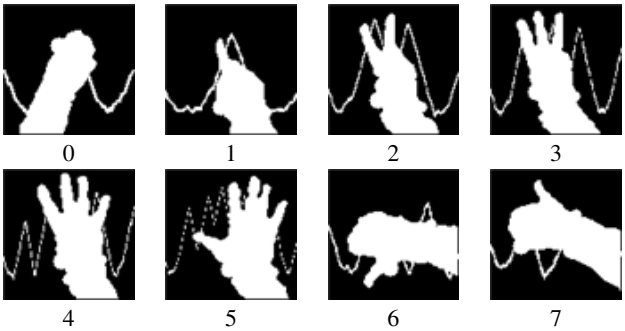


Fig. 6. 8 contour-based hand features.

otherwise. If the measurement likelihood of gesture A is x_A, then $(1-x_A)$ is the probability that the recognition for A can be false. From that, $(1-x_A)^{\gamma_A}$ is the probability of γ_A consecutive false results. If the result of $(1-x_A)^{\gamma_A}$ is less than 5% (0.05), then it can be said that the data have been obtained within a confidence interval (1.96σ, $P = 0.05$) of the 95% confidence level. For example, if the measurement likelihood of gesture A is 80% successively, the recognition failure rate of gesture A might be 20%. The result rate of recognition failure of two consecutive observations is 4% (0.04) and 4% is beyond the 95% confidence interval ($P = 0.05$), so γ of gesture A is 2. At that time, the instance of gesture A is created and vice versa. The likelihood interval-counter using β likelihood distribution can be represented as follows:

$$\gamma_\beta = \min\{\gamma \in I | \prod_{i=1}^{n} (1 - x_{\text{gest}}) \leq P\}, \quad (3)$$

where $P = 0.05 = 1 - 95\%$ confidence level.

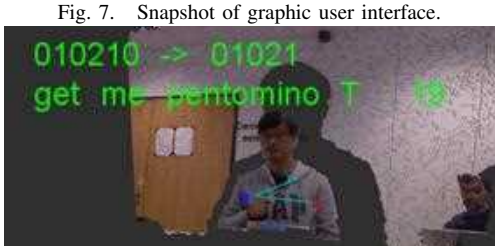
B. Word correction

Additionally, applying a spelling correction method from word corpus [29], the HRI system inferred more robust sentence results out of 24 human commands. The method selects the most likely spelling correction (c) for word (w) out of all possible *candidate* corrections. The most likely correction is the one that maximizes the probability that c is the intended correction, given the original word: $\arg\max_{c \in candidates} P(c|w)$. By Bayes' Theorem this is equivalent to: $\arg\max_{c \in candidates} P(c)P(w|c)/P(w)$. Since $P(w)$ is the same for every possible candidate c, it can be factored out by marginalization, giving: $\arg\max_{c \in candidates} P(c)P(w|c)$, where $P(c)$ is language model that is the probability of c as a word of English text and $P(w|c)$ is the error model that w would be observed when the correct word is c. The method considers both the probability of c and the probability of the change from c to w anyway, so it is cleaner to separate the two factors explicitly. When the *Solver* has information for the human or needs to ask a task on robot initiative, *Multimodal expression* provides two modes: textual messages on screen and speech via ROS *sound_play* node.

VI. Evaluation

Figure 7 shows the graphical user interface. The left bottom panel shows a segmented color image of upper body, while the left middle panel shows a segmented color image of the hand. This hand image and corresponding depth image are fed to CNNs, and the networks produce the loss of

Fig. 7. Snapshot of graphic user interface.



(a) correctly inferred

(b) badly inferred

Fig. 8. Sentence inference results.

prediction. The depth image is also fed to the contour-based recognition module, and the upper left panel shows its result including prediction, the center point of the hand (drawn by a rounded rectangle), and start and end points of the hand (drawn by circles). The right panel shows the detected human skeleton and the current recognized word over the head.

Figure 8 shows the results of sentence inference. When the human raises a hand, the *Sentence Inference* module starts to concatenate the recognition. Then the module makes inference of a sentence when the human lowers the hand. Figure 8 (a) shows a case where the sentence was correctly inferred. The result is also spoken by a speaker. If there is no inferred result, the *Sentence Inference* module asks the human to repeat the gestures, as shown in Fig. 8 (b).

A. Evaluation metrics

To evaluate the interaction and cooperation challenges, both the robot and the human may ask the partner to fetch a P5 or to assemble a given P5 in the puzzle. The objective involves two metrics:

- Metric 1: ability of the robot to recognize gesture commands provided by the human. There are three *Level 1* commands, which are composed of single or simple gestures: *Yes*, *No* and *Get next pentomino*. The other 21 commands are *Level 2*;
- Metric 2: ability of the robot to proactively ask the human to perform certain tasks such as "Fetch next P5" with its identification, "Place the P5 in the Puzzle" with the indication of its position or "Remove P5 from the puzzle" (to make it possible). The interaction initiative for each move is taken alternately by the human and the robot. For robot to human communication the sound speaker and GUI are used;

Figure 9 shows the environment of the scenario.Initially, a challenge host member distributes a set of P5s over the

Overall environment

Assembly station region

Fig. 9. Challenge environments.

(a)

(b)

Fig. 10. Examples of human robot interaction.

pick up station area. Then a human team mate and a mobile robot with a manipulator tries to fetch and assemble the P5s into a fixed squared region in the assembly station area. A human team mate gives a command with a sequence of hand gestures, and receives information through a graphic interface and an audio message, as show in Fig. 10 (a) and (b), respectively.

The evaluation matrix and scores are measured by an external evaluation board which consists of renowned independent experts and are summarized in Table III, where target (T) and baseline (B) are expected points and minimum points set in advance and achievement (A) is acquired points during evaluation. The difficulty points (D) are weighted points decided by challenge board of experts. Improvement (I) is defined by $I = |(A-B)|/|(T-B)|$ and awarded points

(P) $P = I * D$. During the evaluation, all commands including 7 *level 1* and 40 *level 2* were correctly recognized.

	Metric 1	Metric 2	Score
Target	60	20	
Baseline	0	0	
Achievement	87	20	
Improvement	1.45	1	
Deficulty points	3.5	3.5	
Awarded points	5.075	3.5	4.2875

TABLE III
CHALLENGE EVALUATION SCORES

VII. CONCLUSION

This paper describes the human-robot collaboration system developed for the robotics challenges. The system is designed to recognize one-hand gestures and infer a sentence that is a command provided by the human team mate, by integrating several topics such as an interaction graph, hand dataset, human detection, image segmentation, two kinds of feature extractions, gesture recognition, confidence-based dynamic verification, word correction, and multi-modal expression. During the evaluation by the external evaluation board, the received score of *Metric 1* of the Table III is higher than the target score, as more attempts were made than initially planned for the evaluation run.

ACKNOWLEDGMENT

This work was supported by the EuRoC Project under Grant no. 608849 and by National Funds through the FCT - Foundation for Science and Technology, in the context of the project UID/CEC/00127/2013.

REFERENCES

[1] R. Bischoff and T. Guhl, "The strategic research agenda for robotics in europe [industrial activities]," *IEEE Robotics & Automation Magazine*, vol. 1, no. 17, pp. 15–16, 2010.

[2] M. Stenmark, J. Malec, K. Nilsson, and A. Robertsson, "On distributed knowledge bases for robotized small-batch assembly," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 2, pp. 519–528, 2015.

[3] G. Pezzulo, "The "interaction engine": a common pragmatic competence across linguistic and nonlinguistic interactions," *IEEE Transactions on Autonomous Mental Development*, vol. 4, no. 2, pp. 105–123, 2012.

[4] B. Gleeson, K. MacLean, A. Haddadi, E. Croft, and J. Alcazar, "Gestures for industry: intuitive human-robot communication from human observation," in *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*. IEEE Press, 2013, pp. 349–356.

[5] B. Siciliano, F. Caccavale, E. Zwicker, M. Achtelik, N. Mansard, C. Borst, M. Achtelik, N. O. Jepsen, R. Awad, and R. Bischoff, "Euroc-the challenge initiative for european robotics," in *ISR/Robotik 2014; 41st International Symposium on Robotics; Proceedings of*. VDE, 2014, pp. 1–7.

[6] E. Pedrosa, N. Lau, A. Pereira, and B. Cunha, "A skill-based architecture for pick and place manipulation tasks," in *Portuguese Conference on Artificial Intelligence*. Springer International Publishing, 2015, pp. 457–468.

[7] F. Amaral, E. Pedrosa, G. H. Lim, N. Shafii, A. Pereira, J. L. Azevedo, B. Cunha, L. P. Reis, S. Badini, and N. Lau, "Skill-based anytime agent architecture for logistics and manipulation tasks euroc challenge 2, stage ii âĂŞ realistic labs: Benchmarking," in *Autonomous Robot Systems and Competitions (ICARSC), 2017 International Conference on*. IEEE, 2017, accepted for publication.

[8] U. G. Mangai, S. Samanta, S. Das, and P. R. Chowdhury, "A survey of decision fusion and feature fusion strategies for pattern classification," *IETE Technical review*, vol. 27, no. 4, pp. 293–307, 2010.

[9] L. D. Riek, T.-C. Rabinowitch, P. Bremner, A. G. Pipe, M. Fraser, and P. Robinson, "Cooperative gestures: Effective signaling for humanoid robots," in *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*. IEEE, 2010, pp. 61–68.

[10] G. Canal, S. Escalera, and C. Angulo, "A real-time human-robot interaction system based on gestures for assistive scenarios," *Computer Vision and Image Understanding*, vol. 149, pp. 65–77, 2016.

[11] J. Gui, D. Tao, Z. Sun, Y. Luo, X. You, and Y. Y. Tang, "Group sparse multiview patch alignment framework with view consistency for image classification," *IEEE transactions on image processing*, vol. 23, no. 7, pp. 3126–3137, 2014.

[12] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *arXiv preprint arXiv:1405.3531*, 2014.

[13] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, "Robust part-based hand gesture recognition using kinect sensor," *IEEE transactions on multimedia*, vol. 15, no. 5, pp. 1110–1120, 2013.

[14] J. Yang, J.-y. Yang, D. Zhang, and J.-f. Lu, "Feature fusion: parallel strategy vs. serial strategy," *Pattern Recognition*, vol. 36, no. 6, pp. 1369–1381, 2003.

[15] I. Misra, C. L. Zitnick, and M. Hebert, "Unsupervised learning using sequential verification for action recognition," *arXiv preprint arXiv:1603.08561*, 2016.

[16] D. Holz, A. Topalidou-Kyniazopoulou, J. Stückler, and S. Behnke, "Real-time object detection, localization and verification for fast robotic depalletizing," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 1459–1466.

[17] T. Yeh, J. J. Lee, and T. Darrell, "Fast concurrent object localization and recognition," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 280–287.

[18] R. Rusu and S. Cousins, "3D is here: Point cloud library (PCL)," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 2011, pp. 1–4.

[19] M. Oliveira, L. S. Lopes, G. H. Lim, S. H. Kasaei, A. M. Tomé, and A. Chauhan, "3d object perception and perceptual learning in the race project," *Robotics and Autonomous Systems*, vol. 75, pp. 614–626, 2016.

[20] M. A. Goodrich and A. C. Schultz, "Human-robot interaction: a survey," *Foundations and trends in human-computer interaction*, vol. 1, no. 3, pp. 203–275, 2007.

[21] S. Ivaldi, S. M. Anzalone, W. Rousseau, O. Sigaud, and M. Chetouani, "Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement," *Frontiers in neurorobotics*, vol. 8, 2014.

[22] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[23] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th annual international conference on machine learning*. ACM, 2009, pp. 609–616.

[24] Y. Li, "Hand gesture recognition using kinect," in *2012 IEEE International Conference on Computer Science and Automation Engineering*. IEEE, 2012, pp. 196–199.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.

[26] J. Nagi, F. Ducatelle, G. A. Di Caro, D. Cireşan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. M. Gambardella, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *Signal and Image Processing Applications (ICSIPA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 342–347.

[27] A. Bhattachayya, "On a measure of divergence between two statistical population defined by their population distributions," *Bulletin Calcutta Mathematical Society*, vol. 35, pp. 99–109, 1943.

[28] G. H. Lim and I. H. Suh, "Robust robot knowledge instantiation for intelligent service robots," *Intelligent Service Robotics*, vol. 3, no. 2, pp. 115–123, 2010.

[29] P. Norvig, "Natural language corpus data," *Beautiful Data*, pp. 219–242, 2009.