The Knowledge Engineering Review, Vol. 32, e8, 1–9. © Cambridge University Press, 2017
doi:10.1017/S0269888916000321

Inter-humanoid robot interaction with emphasis on detection: a comparison study

TAHER ABBAS SHANGARI[1†], VIDA SHAMS[1†], BITA AZARI[1], FARAZ SHAMSHIRDAR[1], JACKY BALTES[2] and SOROUSH SADEGHNEJAD[1]

[1]*Bio-Inspired System Design Lab, Amirkabir University of Technology (Tehran Polytechnic), No. 424, Hafez Avenue, PO Box 15875-4413, Tehran, Iran*
e-mail: abbasitaher@gmail.com; vidashams1991@gmail.com; bita.az94@gmail.com; faraz.shamshirdar@gmail.com; s.sadeghnejad@aut.ac.ir;
[2]*Autonomous Agents Laboratory, University of Manitoba, Winnipeg, Canada, R3T 2N2*
e-mail: jacky@cs.umanitoba.ca

Abstract

Robot Interaction has always been a challenge in collaborative robotics. In tasks comprising Inter-Robot Interaction, robot detection is very often needed. We explore humanoid robots detection because, humanoid robots can be useful in many scenarios, and everything from helping elderly people live in their own homes to responding to disasters. Cameras are chosen because they are reach and cheap sensors, and there are lots of mature two-dimensional (2D) and 3D computer vision libraries which facilitate Image analysis. To tackle humanoid robot detection effectively, we collected a data set of various humanoid robots with different sizes in different environments. Afterward, we tested the well-known cascade classifier in combination with several image descriptors like Histograms of Oriented Gradients (HOG), Local Binary Patterns (LBP), etc. on this data set. Among the feature sets, Haar-like has the highest accuracy, LBP the highest recall, and HOG the highest precision. Considering Inter-Robot Interaction, it is evident that false positives are less troublesome than false negatives, thus LBP is more useful than the others.

1 Introduction

During the past decades, a considerable amount of literature has been published on cooperative robotics (Cao *et al*., 1997). Multi-robot systems are becoming important as a result of the increasing number of industrial, service, and exploration robots in current use. Moreover, the use of multiple robots working in coordination to execute different types of tasks can bring several advantages over a single robot solution, such as simplicity in robot design, better performance, increased fault tolerance, spatially distributed sensing, and actuation. While many of these tasks can be executed by a single robot, multi-robot systems can do these tasks more efficiently, using simpler robots (Cao *et al*., 1997).

Thus, collaborative robot behaviors will be of an important aspect in scientific and industrial fields such as exploration in ground, space, and underwater, surveillance, assistance, manipulation, assembly of objects in industrial environments, and autonomous rescue operations. Additionally, individual robots need to interact and, in some cases, collaborate with other robots (Ruiz del Solar *et al*., 2010). Looking a bit further ahead, one can easily imagine personal robots going to shop for the family. The user might take the robot to the local mall, where the robot would get authenticated and accepted. The robot can obtain information on goods and their locations, fill the cart (checking the list prepared by smart appliances in the

† These authors contributed equally to this work.

https://doi.org/10.1017/S0269888916000321 Published online by Cambridge University Press

house), wait in the queues etc.—while the owner makes time for more amusing activities (Chaimowicz *et al.*, 2001).

In order to detect objects, we need to record their information by some sensors. Most often, these sensors are classified into active and passive ones. Active sensors provide their own energy for illumination. Active sensors' possibility to measure distance and speed of the objects and the fact that they work well, in bad weather or poor illumination conditions is their main advantage (Discant *et al.*, 2007). Besides the interference problems, other limitations of the active sensors are difficulties in interpreting the output signal returned by these sensors and the acquisition price. Take sonar as a well-known example of active sensors, it transmit sonar waves and detect objects' distance and their speed together. But sonar malfunctions in bad weather conditions. The speed of sound waves varies according to environmental conditions like temperature and pressure.

On the other hand, passive detection can only work when the natural energy is available. Thermal infrared sensors can work in day or night as long as the amount of energy is large enough to be recorded by the sensor's receiver. Visible spectrum camera and the infrared camera are the most two commonly used passive sensors. These sensors most often are called vision-based because they register images.

The infrared sensors have the ability to measure the temperature and because they are independent of the light source they can register the same or almost the same images, even it is day or night. Sensitivity of infrared sensors to weather conditions (rain, fog) remains an issue. Another issue of this type of sensors is that infrared image processing is a relatively new domain for obstacle detection (Discant *et al.*, 2007).

Visible spectrum cameras generally do not emit any signals. Captured images by cameras are very rich in contents, easy to interpret, and able to provide depth information by using stereo images. Additionally, they are not expensive, and also because they are not active, they do not cause interference problems with the environment. These reasons might explain why the most attention in research for object detection was concentrated in this direction, and also why we use this type of sensor (Discant *et al.*, 2007).

The rest of the document is organized as follows. In the next section, we discuss why robot detection is important. What the challenges of robot detection are, and why we chose the human (or humanoid) robots for our study (Gerndt *et al.*, 2015; Baltes *et al.*, 2016). In Section 3, it is explained that how we prepared our data set of humanoid robots, then some advantages and disadvantages of some of the off-the-shelf human detection methods are provided. In the following, we provided an answer that why we used these methods for humanoid robot detection. In Section 4, the experimental results are presented, and finally, we provide a summary of the work and directions for future work (Table 1).

Table 1 Advantages and disadvantages of active and passive sensors

Sensors	Advantages	Disadvantages
Passive		
Visible spectrum camera	Very rich in contents Easy to interpretLow price No interference problems with the environment	Not well suited for darkness conditions
Infrared camera	Ability to measure the temperature Independent of the light source	Sensitive to weather conditions
Active		
Sonar	Possibility to measure distance and speed of the target objects	Difficulties in interpreting the output signal returned by themselves Acquisition price Sensitive to weather conditions
Laser	Having a high accuracy both in lateral and longitudinal direction	Acquisition price
Radar	Images can be acquired day or night Can operate in different environmental conditions without any strong limitations	High price

https://doi.org/10.1017/S0269888916000321 Published online by Cambridge University Press

2 Detection

2.1 Robot detection

We say that on one hand, robotic systems are getting more and more sophisticated, and the range of their use in a various environment is increasing, so in interaction with other robots, they have to track, recognize, and localize them. On the other hand, detection is the primary subsystem of high-level tasks like tracking, recognition, localization, etc. and the accuracy of these systems, typically, depends on the complexity of detection, so an accurate detection allows for a corresponding accuracy of the other (Moeslund and Granum, 2001). Thus, it is highly important to develop accurate and robust robot detection methods by which, enable robots to perform the aforementioned high-level tasks. Robot detection is not only useful in interaction (cooperation) but in observing, ignoring, or even competing with each other. Moreover, there is a complex variety of other types of interactions that will arise among different robots as well as between themselves and humans (Stone, 1998).

However, the development of such methods is a complex task because of the changing conditions of real-world scenarios (e.g. variable illumination and/or cluttered backgrounds), as well as the varying appearance of robots that depends on their relative position in relation to the camera, which is especially important in the case of humanoid and other legged robots. An additional challenge to be taken into account is the currently limited processing power of most robots, which imposes some restrictions on the methodologies that can be used to solve these problems (Ruiz del Solar *et al*., 2010). Even considering a single robot class from a single viewpoint (i.e. the same side of the robot is always facing the camera), detecting them remains a daunting unsolved problem (Schneiderman and Kanade, 2000; Viola and Jones, 2004; Dalal and Triggs, 2005; Zaier, 2012). Achieved good results for faces, cars, and pedestrians. The main issue of these researches is their high false positive rates, when a high recall is required (Coates and Ng, 2010).

As research progresses, there is increasing evidence that the developed prototypes of the robots of tomorrow tend to resemble humans and possess abilities of a human being, including a capacity to walk, speak, type, make decisions, help elderly people live in their own homes, and interact with the real and virtual worlds. Looking a bit further ahead, it is expected that humanoid robots will change the way we interact with machines, and will have the ability to blend perfectly into an environment already designed for humans (Zaier, 2012). In addition to the above mentioned reasons, seeing Defense Advanced Research Projects Agency (DARPA) investing in the research foundation for this field (Guizzo and Ackerman, 2012), motivated us to focus on developing automatic, fast, and accurate methods for humanoid robots detection, in different environments. To address this challenge, the authors present a multi view humanoid robot detection system based on several feature descriptors and classifiers. The successful application of the proposed methodology on our humanoid robot data set showcases its strengths.

2.2 Human detection

In order to be applicable in the real-world tasks, a humanoid robot detection system needs a robust combination of classifier and images descriptor which allows the humanoid robot to be discriminated from other objects with high hit rate and low false alarm. But developing such a system, is challenging due to the humanoid robot's variable appearance and their wide range of poses. Thus, providing a suitable solution which detects humanoid robots efficiently, a careful study has to be done on classifiers, images descriptors, their advantages and disadvantages.

Although humanoid robot is not realistically human-like in appearance and is readily perceived as a robot by human interactants, however, it possess some human-like features, which are usually stylized, simplified or cartoon-like versions of the human equivalents, including some or all of the following: a head, facial features, eyes, ears, eyebrows, arms, hands and legs (https://interestingfactsforkids.wordpress.com/2014/11/16/humanoid-robots/,ed.; Alves *et al*., 2008). In addition to the above mentioned similarities between a humanoid robot and a human, lack of enough researches about humanoid robot detection drove us toward reviewing the image descriptor–classifier combinations used for human detection.

https://doi.org/10.1017/S0269888916000321 Published online by Cambridge University Press

In case of image descriptors, working with image intensities makes feature calculation computationally expensive, unfortunately Haar-like features which are digital image features are based on image intensities. A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums. This difference is then used to categorize subsections of an image. A large number of Haar-like features are necessary to describe an object with sufficient accuracy. For human detection applications, both Histograms of Oriented Gradients (HOG) and Local Binary Patterns (LBP) features outperforms the performance of Haar wavelet feature, which reflects the fact that the intensity pattern of human face is simple than that of human body (Gerndt *et al.*, 2015).

Another feature set, which is popular for object detection, is Histograms of Oriented Gradients. In this feature set, the gradient information in local cells are collected into histograms using trilinear interpolation, and then overlapping blocks composed of neighboring cells are normalized. Interpolation, local normalization and histogram binning make the representation robust to changes in lighting conditions and small variations in pose (Ma *et al.*, 2011). The HOG/scale-invariant feature transform representation (HOG/SIFT) has several advantages. It captures edge or gradient structure that is very characteristic of local shape, and it does so in a local representation with an easily controllable degree of invariance to local geometric and photometric transformations: translations or rotations make little difference if they are much smaller that the local spatial or orientation bin size (Walk *et al.*, 2010).

Though, many experiments have showed that the HOG feature is effective in representation of pedestrian's local shape information, it is still coarse. The HOG features describe the intensity distribution of a pixel or a local region. Because, in real applications, the background is generally complex, and produces some noise points which eventually, decreases the detection performance. The LBP feature can solve these problems well (Dalal and Triggs, 2005). Some examples of combining classifiers with images descriptors in order to detect human are brought in (Gan and Cheng, 2011). 16 state-of-the-art pedestrian detectors are evaluated. The goal was to choose a diverse set of detectors that were both representative of various lines of research and most promising in terms of originally reported performance (Table 2).

3 Our contribution

3.1 Data set

To train and test Haar-Cascade, HOG-cascade, and LBP-cascade combinations, we need to have a data set, containing lots of positive and negative images. Since we could not find a standard data set about humanoid robots, we gathered 100 images of humanoid robot. Because 100 images are not enough to train a classifier well, we randomly scaled or rotated each one these images and created 700 images. In all, 500 images are used for training and 200 are used for testing. All the samples which are created by opencv_createsamples are cropped to standard size of 40×80 pixels in our experiments. The negative images are randomly sampled from the 6000 negative images. Each of these positive examples includes at least one humanoid robot and varies in posture, illumination and background, etc. We evaluated combinations of cascade classifier with three types of feature sets: HOG, LBP, and Haar-like features. Obviously different settings will result in different output. It is worth mentioning that we utilized opencv_traincascade utility to train our cascade classifier. To use opencv_traincascade application, its command line arguments have to be set properly. Below, we discuss the most effective parameters, followed by a short discussion of the future works. Number of positive/negative images, number of stages, acceptance ratio break value, type of features, Haar-like feature parameters, sample width, and sample height are some of the arguments which have to be set carefully.

3.1.1 Number of positive/negative images

Number of positive/negative samples used in training for every classifier stage.

https://doi.org/10.1017/S0269888916000321 Published online by Cambridge University Press

https://doi.org/10.1017/S0269888916000321 Published online by Cambridge University Press

Table 2 An overview of pedestrian detectors (Gan and Cheng, 2011)

	Features							Learning			Detection Details					Implementation		
	Gradient hist.	Gradients	Gray scale	Color	Texture	Self-similarity	Motion	Classifier	Feature learn	Part based	Non-maximum suppression	Model height (inch pixels)	Scales per octaves	Frames per second (fpr)	Log-average miss rate (%)	Training data	Original code	Full image evaluation
VJ (Dollar *et al*., 2012)			✓					AdaBoost			MS	96	~14	0.447	95	INRIA		
Shapelet (Viola and Jones, 2004)		✓						AdaBoost	✓		MS	96	~14	0.051	91	INRIA		
PoseInv (Sabzmeydani and Mori, 2007)	✓							AdaBoost			MS	96	~18	0.474	86	INRIA	✓	
LatSVM-V1 (Lin and Davis, 2008)	✓							Latent SVM		✓	PM	80	10	0.392	80	PASCAL	✓	✓
FtrMine (Felzenszwalb *et al*., 2008)	✓	✓	✓	✓				AdaBoost	✓		PM	100	4	0.08	74	INRIA	✓	
HikSVM (Dollár *et al*., 2007)	✓							HIK SVM			MS	96	8	0.185	73	INRIA	✓	
HOG (Maji *et al*., 2008)	✓							Linear SVM			MS	96	~14	0.239	68	INRIA	✓	
MultiFtr (Dalal and Triggs, 2005)	✓		✓					AdaBoost			MS	96	~14	0.072	68	INRIA	✓	✓
HogLbp (Wojek and Schiele, 2008)	✓				✓			Linear SVM			MS	96	14	0.062	68	INRIA	✓	✓
LatSvm-V2 (Wang *et al*., 2009)	✓							Latent SVM		✓	PM	96	10	0.629	63	INRIA	✓	✓
Pls (Felzenszwalb *et al*., 2010)	✓			✓	✓			PLS + QDA	✓		PM*	96	~10	0.018	62	INRIA	✓	✓
MultiFtr + CSS (Schwartz *et al*., 2009)	✓					✓		Linear SVM			MS	96	~14	0.027	61	TUD-MP	✓	✓
FeatSynth (Walk, *et al*., 2010)	✓				✓			Linear SVM	✓	✓	–	96	–	–	60	INRIA	✓	✓
FPDW (Bar-Hillel *et al*., 2010)	✓	✓	✓	✓				AdaBoost			PM*	100	10	6.492	57	INRIA	✓	✓
CHNFTRS (Dollár *et al*., 2010)	✓	✓	✓	✓				AdaBoost			PM*	100	10	1.183	56	INRIA	✓	✓
MultiFtr + Motion (Dollár, *et al*., 2009)	✓					✓	✓	Linear SVM			MS	96	~14	0.02	51	TUD-MP	✓	✓

HIK = histogram intersection kernel; SVM = support vector machine; PLS = partial Least Squares; QDA = quadratic discriminant analysis; MS = multi scale; PM = pairwise max; PASCAL = programming language; TUD-MP = TUD-MotionPairs.

3.1.2 Number of stages
Number of cascade stages to be trained.

3.1.3 Acceptance ratio break value
This argument is used to determine how precise your model should keep learning and when to stop. A good guideline is to train not further than 10*e*–5, to ensure the model does not overtrain on your training data. By default, this value is set to −1 to disable this feature.

3.1.4 Type of features
Type of features: Haar—Haar-like features, LBP.

3.1.5 Haar-like feature parameters
Selects the type of Haar features set used in training. BASIC uses only upright features, while ALL uses the full set of upright and 45° rotated feature set.

3.1.6 Sample width/height
Size of training samples (in pixels) must have exactly the same values as used during training samples creation (opencv_createsamples utility).

In this research, the above mentioned parameters are set as follows: number of positive images is set to 100, number of negative images is set to 500, number of stages is set to 12, and sample width/height is set to 40×80.

Since the test ground-truth (GT) labels are not available, we split the data set into train and validation sets. Testing is performed in two modes. In the first mode, in order to calculate the recall rate, we tested the proposed solutions with 100 positive samples. In the second mode, we provided 100 negative samples, and calculated true negative and false positive rates of the solutions. As an example, Figures 1 and 2, respectively, illustrate output of the proposed solutions on two negative images and five positive images.

For each GT object, we find the best humanoid robot proposal. We say that a GT instance has been recalled if the best proposal exceeds 50% of the humanoid robot (Figure 3).

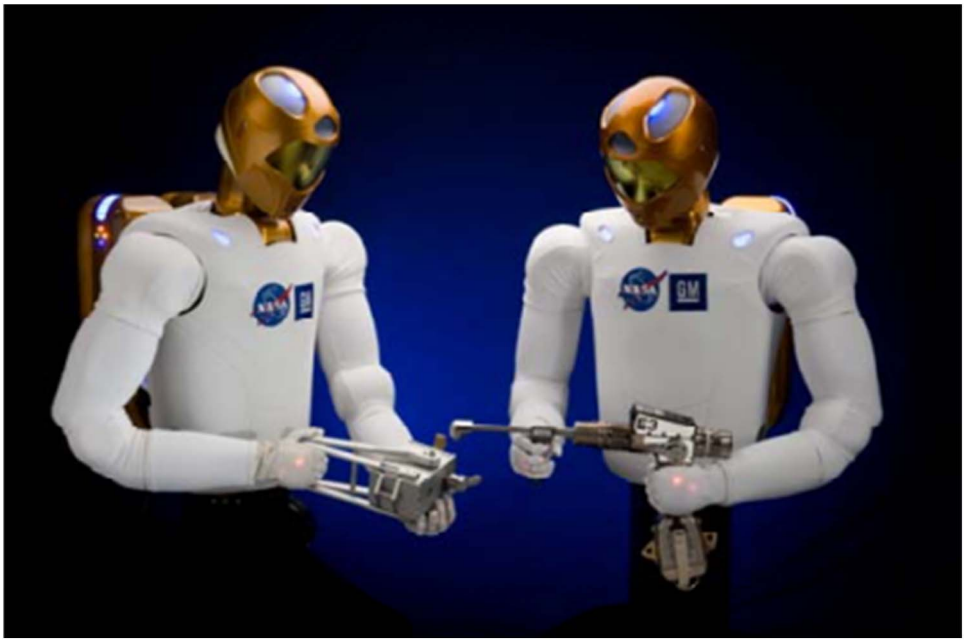


Figure 1 Second generation Robonauts are collaborating (`http://www.nasa.gov/images/content/421731main_jsc2009e155295.jpg`, ed)

https://doi.org/10.1017/S0269888916000321 Published online by Cambridge University Press

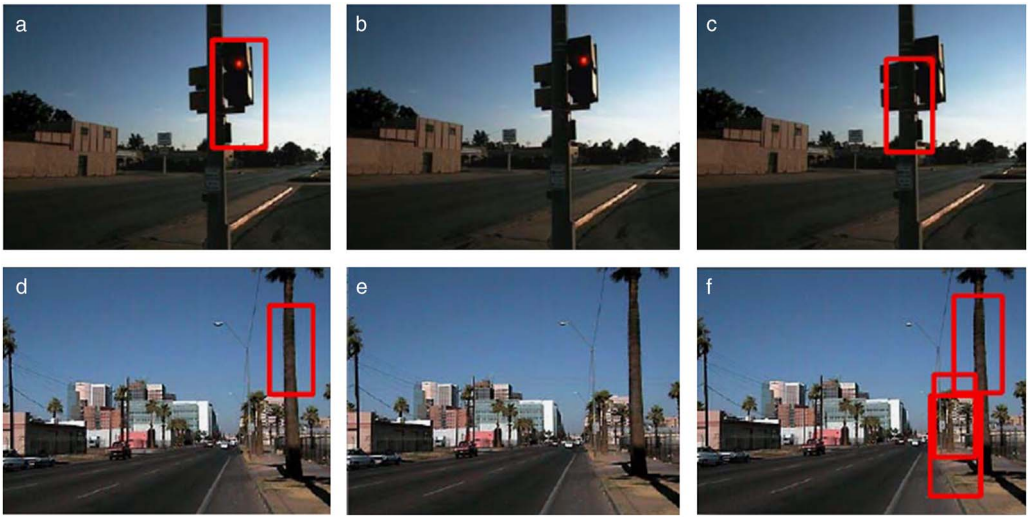


Figure 2 (a) and (d) are Local Binary Patterns output, (b) and (e) are Histograms of Oriented Gradients output, (c) and (f) are Haar's output. These images are presented to show result of the image descriptors on images which do not contain a robot. The original images are taken from French Institute for Research in Computer Science and Automation Person image data set (Walk *et al.*, 2010)

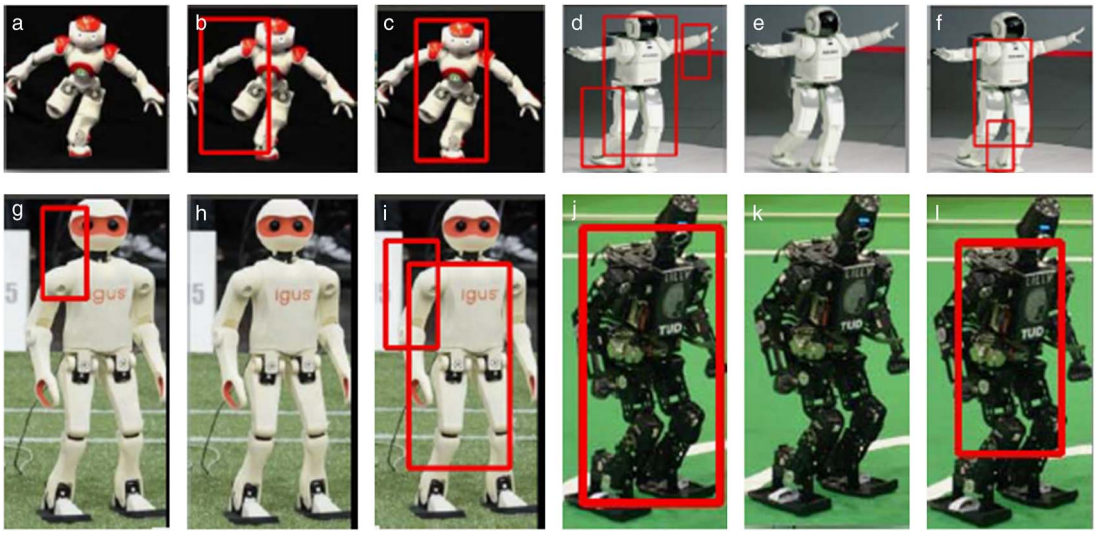


Figure 3 (a), (d), (g), and (j) are Haar's output, (b), (e), (h), (k) and (e) are Histograms of Oriented Gradients output, (c), (f), (i), and (l) are Local Binary Patterns output. In these images, it is evident that LBP has the highest rate of true positives in comparison with other feature sets. Images are taken from `ftp://ftp.inrialpes.fr/pub/lear/douze/data/`

Table 1 reveals that Haar-like feature set has the highest accuracy, LBP the highest recall, and HOG the highest precision. Considering inter-robot interaction, it is evident that false positives are less troublesome than false negatives, thus recall rate is more important than the other parameters. In order to improve the precision rate of LBP, we are going to combine it with HOG in the future works.

4 Conclusion

As we have shown, recall rate of LBP feature set coupled with a cascade classifier outperforms HOG and Haar. But, LBP suffers from low precision. However, the problem of high false positives which LBP suffers from, can be solved using background subtraction methods, and does not cause a serious problem.

https://doi.org/10.1017/S0269888916000321 Published online by Cambridge University Press

Table 3 Comparison of results obtained from Haar-like, LBP and HOG descriptors

Feature sets	True positive (%)	True negative (%)	Recall (%)	Precision (%)	Accuracy (%)
Haar-like	88	80	88	82	84
LBP	90	39	90	59.60	65
HOG	45	98	45	96	71.50

In the future works, we are going to combine HOG and LBP feature sets in order to achieve higher recall and precision rate. Eventually, we hope to extend our work to more complex tasks such as humanoid robot tracking, identification, etc. We would also like to explore additional combination of feature sets and classifiers, and also make our data set diverse (Table 3).

Acknowledgement

The authors would like to thank the Bio-Inspired System Design Laboratory (BInSDeLa) of Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran and also Autonomous Agents Laboratory, University of Manitoba, Winnipeg, Canada. This research has been done in collaboration of both universities under the corporation protocol, started from 2014.

References

Alves Jr, A.A., Andrade Filho, L.M., Barbosa, A.F., Bediaga, I., Cernicchiaro, G., Guerrer, G., Lima Jr, H.P., Machado, A.A., Magnin, J., Marujo, F. and De Miranda, J.M., 2008. The LHCb detector at the LHC. *Journal of instrumentation*, 3(08), p.S08005.

Baltes, J., Tu, K.-Y., Sadeghnejad, S. & Anderson, J. 2016. HuroCup: competition for multi-event humanoid robot athletes. *The Knowledge Engineering Review*, 1–14.

Bar-Hillel, A., Levi, D., Krupka, E. & Goldberg, C. 2010. Part-based feature synthesis for human detection. In *European Conference on Computer Vision*, 127–142. Springer Berlin Heidelberg.

Cao, Y. U., Fukunaga, A. S. & Kahng, A. 1997. Cooperative mobile robotics: antecedents and directions. *Autonomous Robots* **4**, 7–27.

Chaimowicz, L., Sugar, T., Kumar, V. & Campos, M. F. 2001. An architecture for tightly coupled multi-robot cooperation. In *Proceedings 2001 ICRA, IEEE International Conference on Robotics and Automation, 2001*, 2992–2997.

Coates, A. & Ng, A. Y. 2010. Multi-camera object detection for robotics, In *IEEE International Conference on Robotics and Automation (ICRA), 2010*, 412–419.

Dalal, N. & Triggs, B. 2005. Histograms of oriented gradients for human detection, In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, CVPR 2005*, 886–893.

Discant, A., Rogozan, A., Rusu, C. & Bensrhair, A. 2007. Sensors for obstacle detection-a survey, In *30th International Spring Seminar on Electronics Technology*, 100–105.

Dollár, P., Belongie, S. & Perona, P. 2010. The Fastest Pedestrian Detector in the West. *BMVC* **2**(3).

Dollar, P., Wojek, C., Schiele, B. & Perona, P. 2012. Pedestrian detection: An evaluation of the state of the art. *IEEE transactions on pattern analysis and machine intelligence* **34.4**, 743–761.

Dollár, P., Tu, Z., Perona, P. & Belongie, S. 2009. Integral channel features, 91–1.

Dollár, P., Tu, Z., Tao, H. & Belongie, S. 2007. Feature mining for image classification, In *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07*, 1–8.

Felzenszwalb, P., McAllester, D. & Ramanan, D. 2008. A discriminatively trained, multiscale, deformable part model, In *IEEE Conference on Computer Vision and Pattern Recognition, 2008, CVPR 2008*, 1–8.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D. & Ramanan, D. 2010. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**, 1627–1645.

Gan, G. & Cheng,, J. 2011. Pedestrian detection based on HOG-LBP feature, In *Seventh International Conference on Computational Intelligence and Security (CIS), 2011*, 1184–1187.

Gerndt, R., Seifert, D., Baltes, J. H., Sadeghnejad, S. & Behnke, S. 2015. Humanoid robots in soccer: robots versus humans in RoboCup 2050. *IEEE Robotics & Automation Magazine* **22**, 147–154.

Guizzo, E. & Ackerman, E. 2012. *How rethink robotics built its new Baxter robot worker*. `http://spectrum.ieee.org/robotics/industrial-robots/rethink-robotics-baxter-robot-factory-worker`.

https://doi.org/10.1017/S0269888916000321 Published online by Cambridge University Press

Lin, Z. & Davis, L. S. 2008. A pose-invariant descriptor for human detection and segmentation. In *Computer Vision–ECCV 2008*. Springer, 423–436.

Ma, Y., Chen, X. & Chenm, G. 2011. Pedestrian detection and tracking using HOG and oriented-LBP features. *Network and Parallel Computing*, 176–184.

Maji, S., Berg, A. C. & Malik, J. 2008. Classification using intersection kernel support vector machines is efficient, In *IEEE Conference on Computer Vision and Pattern Recognition, 2008, CVPR 2008*, 1–8.

Moeslund, T. B. & Granum, E. 2001. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding* **81**, 231–268.

Ruiz del Solar, J., Verschae, R., Arenas, M. & Loncomilla, P. 2010. Play Ball!. In *IEEE Robotics & Automation Magazine*, **17**(4), 43–53.

Schneiderman, H. & Kanade, T. 2000. A statistical method for 3D object detection applied to faces and cars, In *Proceedings. IEEE Conference on Computer Vision and Pattern Recognition, 2000*, 746–751.

Sabzmeydani, P. & Mori, G. 2007. Detecting pedestrians by learning shapelet features, In *IEEE Conference on Computer Vision and Pattern Recognition, 2007, CVPR'07*, 1–8.

Schwartz, W. R., Kembhavi, A., Harwood, D. & Davis, L. S. 2009. Human detection using partial least squares analysis, In *IEEE 12th international conference on Computer Vision, 2009* , 24–31.

Stone, P. 1998. *Layered Learning in Multiagent Systems: A Winning Approach to Robotic Soccer*. MIT Press.

Viola, P. & Jones, M. J. 2004. Robust real-time face detection. *International Journal of Computer Vision* **57**, 137–154.

Walk, S., Majer, N., Schindler, K. & Schiele, B. 2010. New features and insights for pedestrian detection, In *IEEE conference on Computer Vision and Pattern Recognition (CVPR), 2010* , 1030–1037.

Wang, X., Han, T. X. & Yan, S. 2009. An HOG-LBP human detector with partial occlusion handling, In *IEEE 12th International Conference on Computer Vision, 2009*, 32–39.

Wojek, C. & Schiele, B. 2008. A Performance Evaluation of Single and Multi-feature People Detection. In *Proceedings of the 30th DAGM symposium on Pattern Recognition*, Rigoll, G. (ed.). Springer-Verlag, Berlin, Heidelberg, 82–91.

Zaier, R. 2012. *The Future of Humanoid Robots – Research and Applications*. InTechCorp.

https://doi.org/10.1017/S0269888916000321 Published online by Cambridge University Press