IEEE COMSOC MMTC E-Letter

Robotics-Derived Requirements for the Internet of Things in the 5G Context

Giorgio Metta
Istituto Italiano di Tecnologia, Genoa, ITALY
giorgio.metta@iit.it

1. Introduction

Recent advances in robotics indicate that a vigorous market of personal or service robot is certainly possible in the near future. There is a clear demographic trend that would make robotics very appealing to a large sector of the consumers, health care and industrial market, in particular, as generic helpers in the household or factory.

The EU Commission has published statistics that illustrate the dramatic shift in the population distribution from now to 2050. In particular, the number of people out of the working age (i.e. >65 years old) will be as much as one third of the population [1]. Therefore, for each person over 65, there will be at most two other individuals to take care of her/him. Barring dramatic societal changes, the overall cost of health care will top 29% of the European GDP [2]. Robotics appears to be the most feasible technology to deploy effective solutions.

Such a pervasive use of robots will require a robust infrastructure to connect to/from the robot sensors as well as to the traditional telecommunication means (Internet, mobile, etc.). In this letter, we analyze the requirements in terms of data types and bandwidth needed to perform typical robotic tasks.

2. Robot helpers

The future robotic helpers are typically envisioned as machines with autonomy, highly sophisticated sensors, complex mechatronics and flexibility to be deployed in the most disparate applications [3]. They will effectively be the next generation of personal devices, with one major difference if compared to current digital technology: robots will physically interact and change the state of the environment by pressing buttons, levers, moving objects, and occupying space.

One important aspect of the use of robots whenever they share space with people is safety. In spite of the use of compliant robots, force control, etc. absolute safety is only obtained when the robot is guaranteed not to touch a person unless the task requires it to do so. Therefore the use of visual sensors is mandatory. Computer vision algorithm for example will recognize and track the interaction with people and foresee contacts avoiding them when dangerous and, on the contrary, controlling them when needed.

Further, robots will need large knowledge bases to deal with the variety of tasks, objects, people that are encountered in everyday life. These will need to be accessible with certain deterministic latencies to guarantee the continuity of the robot action.

Robots will form part of the machine to machine communication (M2M) paradigm being able to team up to solve tasks, to exchange important information about the ongoing state of the operative environment, etc. Specific protocols and data distribution services may be required to make the joint use of robots and telecommunication infrastructure as effective as possible, especially in the 5G context.

Robots will be remotely-operated to various extents and/or partially monitored (human in the loop approach) particularly in the first commercial applications where full autonomy is unlikely (for regulatory as well as technical reasons). They will require highly reliable and efficient connectivity especially with respect to the overall latency.

Finally, computational infrastructure in the cloud is expected to be needed for robots to perform demanding tasks such as learning. In that case training data have to be transferred, elaborated by machine learning techniques (which may take days to complete) and returned to the robot in the form of specific algorithms (or their parameters).

3. A case study: the iCub

The iCub [4] is a humanoid robot developed by the Istituto Italiano di Tecnologia (Italian Institute of Technology, IIT) and supported by a stream of EU projects in the sixth and seventh framework programs. The iCub is about 1m tall (see Figure 1) and designed to resemble a four-years-old child. The robot principal application domain is research and in particular the study of artificial cognitive systems including AI, learning, vision, control, etc.

The iCub is, to the best of our knowledge, the only robot equipped with a full body skin system that provide tactile information in more than 4000 sensing points distributed along the entire body [5]. The iCub mechanical design sports highly dexterous hands (9 degrees of freedom, 22 joints), a controllable visual system that includes vergence and three degrees of

freedom (DoF) in the head, a torso with additional three DoF and legs with special compliant actuators.

From the sensorial point of view, besides the skin system, the iCub has cameras, microphones, force sensors, gyroscopes, accelerometers that are managed by a distributed computation architecture. The robot software system is intrinsically networked. Information is distributed to a cluster, elaborated and finally transformed into control signals to generate the robot behaviors [6].

This data processing paradigm is in a sense ready to be transformed into "cloud computing" provided a reliable infrastructure is available. On top of that, the iCub can function as a data collection device, by providing its sensory data to other devices, to other robots, or to human operators [6].

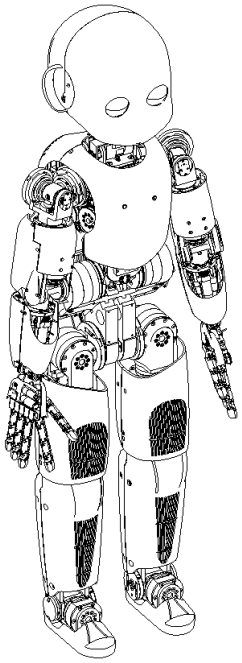


Figure 1: The iCub robot platform.

The iCub is not fully autonomous. Although batteries can be easily fitted, the connectivity with the main computers requires a bandwidth in the range of 1Gbit/s. This accommodates the real-time streaming of two image streams at 30 fps at the resolution of 640x480 (which is the minimum required for most applications). Images are typically sent using Multicast protocols to a set of computers for parallel processing. In a typical laboratory setting, the iCub is tethered. Gbit/s Wi-Fi is now available commercially and in the near future may represent a good substitute for the cable. If we imagine robots working in open/public spaces, then we need to think of something else as for example 5G networks.

4. Typical data stream and processing

The robot data processing can be divided into two main categories: local and remote. Tight real-time control loops are typically executed locally. On the other hand, visual processing cannot be handled locally (because of the computational load/amount of data) and therefore is managed by networked computers (remote). As long as we can rely on the Gbit/s network, it is clear that the distinction is somewhat academic. In the future though, if the robot has to become fully autonomous, then it is also fundamental to consider where to compute, what to compute and how to transmit it given the available bandwidth (e.g. 5G).

The following table summarizes the data rates and estimated bandwidth for all robot sensors and control signals.

Sensor name	Specs	Bandwidth
Cameras	2x, 640x480, 30fps, 8/24bit	147Mbit/s uncompressed
Microphones	2x, 44kHz, 16bit	1.4Mbit/s
F/T sensors	6x, 1kHz, 8bit	48kbit/s
Gyroscopes	12x, 100Hz, 16bit	19.2kbit/s
Tactile sensors	4000x, 50Hz, 8bit	1.6Mbit/s
Control commands	53DoF x 2-4 commands, 100Hz/1kHz, 16bit	3.3Mbit/s (worst case), 170kbit/s (typical)

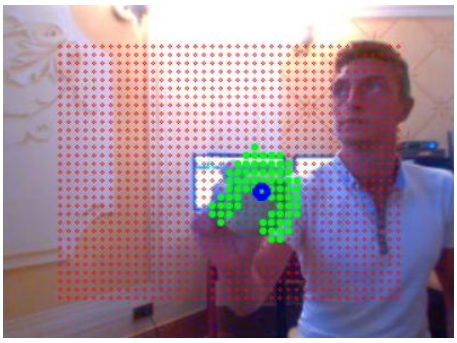
Clearly, image data are the most demanding. It is instructive to briefly discuss the type of processing typically required to extract iconic information from the data streams for a given set of task.

For a helper robot to be useful in the household or factory, it has to be able to move from A to B (independently), fetch objects/tools, use them and possibly go back to A if required by the task. We can list the hypothetical control/processing modules as:

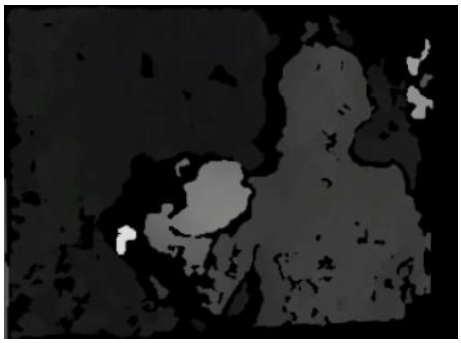
- **Walking controller**: uses vision to do localization and mapping (SLAM) to build a representation of the environment, avoid obstacles;
- **People detector**: uses vision to identify people, their body postures and actions for interaction [7];
- **Object detector/recognition**: uses vision to identify objects (relevant to the task at hand), tools and machines (also potentially relevant to accomplish a task) [8].

All these subsystems require 3D vision obtained from lasers or stereo cameras, motion estimation (optical flow calculation), feature extraction (e.g. SIFTs, HoGs, etc.), and clearly synthetizing a number of controllers that use this information to actually effect proper behaviors. An example of the image processing is shown in Figure 2.

The calculation of the 3D structure of the environment requires camera calibration and then binocular disparity estimation. If this is carried out "on board", then iconic 3D information about surfaces or 3D features can be transmitted outside the robot for further processing (e.g. object identification). SLAM is likely to be done on board in order to generate walking patterns with minimum latencies. Map information can be exchanged with servers to dynamically load/update the maps. A swarm of robots can for example maintain a shared updated map of the environment.



(a)



(b)

Figure 2: Examples of image processing for human-robot interaction: (a) optical flow computation used for object tracking; (b) binocular disparity and 3D structure estimation (see [10]).

Optical flow is computed by processing sequences of images and estimating the pixel-level variations due to motion. Efficient methods are now available. The same information can be used to compress the image stream similarly to the various MPEG or H26xx formats. Motion estimation is useful for action recognition and to predict the movement of objects/people in the scene in order to task the robot consequently.

In object recognition instead, we need to extract local visual features, group them, and further elaborate on them to build classifiers of various sorts. Machine learning has been particularly successful recently on these tasks. Clearly also in this case there is a serious tradeoff between on board computation (feature extraction is demanding) and the bandwidth required to send images to external servers.

It must be said, that the recent development of image processors (GPUs) for the mobile world may come to rescue here [9]. A set of GPUs may in fact be hosted directly on the robot taking care of all the heavy post-processing and subsequently communicating only sporadically to the servers where global knowledge is stored. In this case, objects would become only a set of parameters that can be loaded once (e.g. as function of the task, location, etc.) and maps can be loaded and updated by parts.

More importantly perhaps, latencies have to be controlled and guaranteed. A controller may need a reply from an external server within a given deadline otherwise the robot may need to stop. Also, in human-robot interaction, answers from remote servers need to be fast enough in order not to disrupt the user experience (in the few milliseconds range). This is to be considered in the preparation of the protocols and the hardware infrastructure of the future generation network (e.g. 5G).

4. Conclusion

In this letter, we briefly surveyed the overall requirements in terms of connectivity of a complex robot designed for human-robot interaction believed to become the paradigmatic robot helper of the near future. We imagine a future where swarms of robot helpers will form part of our daily life.

We broadly observed that vision is the most demanding sensory modality in terms of data rates. We also argued that vision is strictly necessary to guarantee the robot safety at all time and that some of the visual information may need to be transmitted outside the robot (either to guarantee the robot's autonomy or for tele-operation).

We do not support a brute force approach (transmit everything) but rather a more parsimonious definition of the computational resource allocation leading to a specialized channel for iconic information in the form of object identities, shapes, surfaces, as well as, actions, people and in general scene maps.

Furthermore, specialized QoS channels may be required if remote controlled operations with safety constraints need to be supported. In that case, maximum message delays need to be specified as well.

We believe that robots will be heavy "users" of the future mobile infrastructure and would be utterly important to prepare the underlying protocols already for this scenario.

Clearly, this paper is highly speculative and high level. It is only claiming that this is a likely scenario and therefore that it is useful to be ready to complement robotics with telecommunication infrastructure. On the other hand, additional resources need to be integrated. We mentioned briefly the need of new low-power visual processing (GPUs) but also efficient batteries that guarantee hours of independent use of the robot. New materials are important for robotics and will certainly constitute a fertile line of research in the near future.

References

[1] http://ec.europa.eu/economy_finance/articles/structural_reforms/2012-05-15_ageing_report_en.htm

[2] http://europa.eu/rapid/press-release_IP-05-322_en.htm

[3] B. Gates. (2007, January 1st, 2007) A robot in every home. Scientific American.

[4] A. Parmiggiani, M. Maggiali, L. Natale, F. Nori, A. Schmitz, N. Tsagarakis, J. Santos-Victor, F. Becchi, G. Sandini, and M. G., "The design of the iCub humanoid robot," International Journal of Humanoid Robotics vol. 9, pp. 1-24, 2011.

[5] P. Maiolino, M. Maggiali, G. Cannata, G. Metta, L. Natale. 2013, A Flexible and Robust Large Scale Capacitive Tactile System for Robots. IEEE Sensors Journal. 13(10) pp.3910-3917.

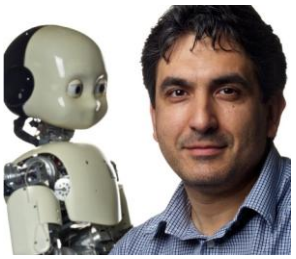
[6] P. Fitzpatrick, G. Metta, L. Natale. 2008, Towards Long-Lived Robot Genes. Robotics and Autonomous Systems. 56(1) pp.29-45.

[7] S.R. Fanello, I. Gori, G. Metta, F. Odone. 2013, Keep It Simple And Sparse: Real-Time Action Recognition. Journal of Machine Learning Research. pp.2617-2640.

[8] I. Gori, U. Pattacini, V. Tikhanoff, G. Metta. Three-Finger Precision Grasp on Incomplete 3D Point Clouds. IEEE International Conference on Robotics and Automation (ICRA2014). Hong Kong, China, May 31-June 7, 2014.

[9] NVIDIA Jetson TK1 development kit website: http://www.nvidia.com/object/jetson-tk1-embedded-dev-kit.html

[10] I. Gori, S.R. Fanello, F. Odone, G. Metta. A Compositional Approach for 3D Arm-Hand Action Recognition. IEEE Workshop on Robot Vision (WoRV). Clearwater, Florida, USA, January 16-18, 2013.



Giorgio Metta is director of the iCub Facility department at the Istituto Italiano di Tecnologia (Italian Institute of Technology, IIT) where he coordinates the development of the iCub robotic platform/project. He holds an MSc cum laude (1994) and PhD (2000) in electronic engineering both from the University of Genoa. From 2001 to 2002 he was postdoctoral associate at the MIT AI-Lab. He was previously with the University of Genoa and since 2012 Professor of Cognitive Robotics at the University of Plymouth (UK). He is deputy director of IIT delegate to the international relations and external funding. In this role he is member of the board of directors of EU Robotics AISBL, the European reference organization for robotics research. Giorgio Metta research activities are in the fields of biologically motivated and humanoid robotics and, in particular, in developing humanoid robots that can adapt and learn from experience. Giorgio Metta is author of approximately 250 scientific publications. He has been working as principal investigator and research scientist in about a dozen international as well as national funded projects.