Human-Centered Computer Interaction System – a Vision-Based Approach

Michal Podpora, Aleksandra Kawala-Janik
Faculty of Electrical Engineering, Automatic Control and Informatics, Opole University of Technology, Poland
45-272 Opole, ul. Sosnkowskiego 31
m.podpora@po.opole.pl, a.kawala-janik@po.opole.pl

Katarzyna Rudnik
Faculty of Production Engineering and Logistics
Opole University of Technology, Poland
45-036 Opole, ul. Luboszycka 5
k.rudnik@po.opole.pl

Abstract— **In this paper authors describe an innovative idea of a vision-based human-centered interaction system. The novelty of the proposed conception involves the implementation of the computer vision as the main information input. The proposed framing enables potential application of the computer vision in role of key input in systems designed for nondeterministic, human-based environment. Presented solution is based on novel image processing method, where an innovative image representation was applied in order to provide a reliable quality of the image and to override existing limitations – such as those occurring during image data transfer between the computer cluster and the system. The idea of the research is based on fact that vision is the most valuable and efficient information source for human beings and therefore the proposed method is based upon its biological origin.**

Keywords—Human-Computer Interaction, Human-Machine Interaction, Autonomous Systems, Autonomous Policies, Expert Systems, Robots, Computer Vision, Distributed Processing, Image Processing

I. INTRODUCTION

Vision is the most valuable and significant information source for Human. It is essential not only for perception of the world, but also for noticing/creating subjective similarities/connections between different objects sharing similar visual properties. Learning/teaching is more efficient if the knowledge/information is given in a graphical way. It makes the whole process more natural and intuitive. Interpersonal communication (exchange of information) is also more efficient if both sides of the interaction have the possibility to visualize the information (e.g. it is easier to repair a car for someone who saw how to do that than for someone who heard how to do that).

In order to bring the human-computer interaction to the same level, the interaction should be performed in a natural way – not only without user-operated input/output devices, but also using the same senses and closest possible cognitive representations of the world, objects, properties, etc. (qualia).

In this paper, a brief summary of research, experience and practice of human-oriented design of a vision-based system framework (developed by the authors) for distributed cognitive systems is presented. Computer Vision (CV) is a massively explored research topic and there are many successful implementations of CV subsystem as a basic information source for a computer/robotic system. There are some systems capable of learning new information (e.g. new visual objects) using only the visual input and voice commands [1]. However, these systems can be taught only a limited number of new objects due to computational limitations of systems implemented in the mobile hardware. These computational limitations can be overridden with the use of distributed systems, however:

- It is currently not possible to embed a truly scalable distributed system into a mobile robotic platform (e.g. due to its size, weight, power-consumption, etc.).
- If the distributed system is accessed remotely from the mobile platform, the connection should support enough bandwidth to transmit video stream.
- If the distributed system is designed to memorize, store and process visual information on many real-world objects, it should have not only proper structures (cognitive system, semantic network, etc.) but first of all it should be able to handle the traffic (the data volume on a video processing distributed system makes the data transmission a bottleneck for the processing algorithms).

There are some implementations of distributed systems for image processing, but they usually apply to deterministic environment and a limited object count (e.g. in industrial applications – on a production line as a quality check). Implementation of a distributed system for human-centered vision-based interaction was, at first, not possible and later not efficient enough to handle a real-world environment and numerous objects.

In [2] and [3], a novel approach was presented, in which a new image representation was proposed to provide a fair quality image and to override existing limitations (transmission of image data between computer cluster nodes as a bottleneck of efficient data processing). The image representation and the distributed vision-based cognitive system framework are described briefly in the following sections.

High quality image data is crucial for any vision-based system, where image processing is being preformed. One of the main issues is that the annotation of large data sets could take a lot time and cause significant delays in image recognition process [3, 9].

II. Image Representation for the Proposed System

The image representation, proposed in [2], is based on Active Vision (AV), where some areas of an image are more significant than the rest of image. In the most common use of AV, only a fragment of an image is being analyzed (e.g. reading ID-plates of cars passing by). Another use of AV involves scanning throughout image (whole image or only a part of an image) – e.g. perception of jamb/doorway for mobile robots. The proposed image representation involves acquiring and transmitting whole image, but the quality of image differs with the distance from a specified coordinates. The proposed compression algorithm is lossless in the neighborhood of the coordinates, offering the best possible quality. The AV-based Point-of-Interest conception with variable threshold was described briefly in [2] and in more detail in [3].

The key features of the proposed image representation are:

- The quality of vision differs spatially – the quality loss of peripheral vision as well as the greatest quality in the fixation point was inspired by human vision and the yellow spot of human eye.
- There is no steep quality change in the whole image.
- The Point-of-Interest (POI) (e.g. the coordinates of the center of the best image quality area) is nominated by inferring algorithm and passed back to image acquisition subsystem in a loop-back.
- The image processing algorithms must handle the lower quality of peripheral vision and support/participate in the POI election procedure.

The proposed image representation and processing procedures, just like any other AV-based algorithm, require feedback information about Region-of-Interest (in this case: Point-of-Interest). Therefore, the Distributed Video-based Cognitive System should have the POI election procedures implemented.

III. Image Representation for the Proposed System

The Distributed Cognitive System (D.C.S.) is intended to be a remote computational resource for mobile robots. In its current implementation a mobile robot can connect as a client and take advantage of the cognitive system to enhance its cognitive abilities. The exemplary mobile robot – without connection to D.C.S. – is capable of driving along the wall and building a basic map basing on motor encoders. However, when the wireless connection to D.C.S. is established, the robot is able to enter autonomous discovery mode and use Computer Vision as information source.

D.C.S. in its current implementation is based upon a Computer Cluster and a simple Semantic Network. One cluster node is dedicated to manage data connection with a robot and to initiate visual data analysis (including image processing and basic inferring) using Functional Distribution. An early version of D.C.S. is described in [3].

In the current version of D.C.S. implementation it is possible to perform basic supervised learning tasks – with the use of keyboard as the teacher's input (feedback) device, while robot navigates autonomously and asks questions about 'interesting' objects (e.g. colorful or moving) in the scene.

It also important to mention that in the current version of D.C.S., a robot is neither programmed to store the information, nor to formulate autonomous compound actions basing on knowledge of the objects. These features require a dedicated cognitive architecture, which has not been fully implemented yet.

IV. Knowledge Acquisition, Storage and Inferring

One of the main problems to solve when designing a cognitive system is knowledge management (building and processing of knowledge as well as inference). Knowledge representation in a form of a Semantic Network (SN) allows submitting the knowledge (visual data) in the form of two-argument relationships between objects (i.e. nodes of the SN). Network's hierarchical representation supports building up the virtual representation of an image. The main advantage of Semantic Networks is their flexibility and lack of space constraints. However, the process of building up a SN requires definitions of both: objects and relations between them, as well as a definition of unambiguous interpretation record. Construction of Semantic Networks for any real-world application of computer vision is therefore a very complex problem. Reducing the problem to a specific area (room, the view, the slot of production) or to limited set of objects and environments is only a partial simplification. Authors propose to use idea of knowledge mining as a method for building up a Semantic Network. Knowledge mining can be characterized by the following formula [4]:

Data + Prior Knowledge + Goal $\rightarrow$ New Knowledge

where "data" constitute encoded knowledge of current picture (or – on the sensory level – pixels of current image), "prior knowledge" are pre-defined objects in the Semantic Network and/or pre-defined network, and the "goal" is the identification of needs related to the creation of knowledge. *A priori* knowledge includes predictable or static background of video input. The attributes of objects can be described by linguistic values. This representation is simple to record and easy to understand by a human.

The Semantic Network, after building it up from the prior knowledge, becomes a new representation of knowledge (new knowledge) of acquired visual information about the scene. Generating new knowledge leads to teaching Semantic Network model - detecting changes in the model.

The essence of the approach is to use data mining methods to create new knowledge based on dynamic video stream data. Data mining, as the main stage of knowledge discovery [5], deals with algorithms, which discover "hidden" (previously unknown and potentially necessary/important) information in the input data [6] and store it as patterns and knowledge models.

Depending on the knowledge representation, there are many algorithms and methods of data mining. Sequence analysis allows recording dynamic video sequences. Static analysis of video frames with an appropriate frequency allows to use other methods such as clustering or association rules.

The frequency can be tuned to extract specific image/objects features or properties.

Data mining algorithms are used for analysis of large volumes of data in data warehouses, but usually not as on-line analysis methods. Therefore, distributed execution of the tasks has also been considered for additional calculations acceleration.

Understanding the vision allows to go one step further – to infer basing on the created knowledge. Representation of Semantic Networks in the form of conditional rule base makes it possible to infer about the content of the transmitted image. Semantic Network's outputs can also be considered as the input of the decision-making system. By that means, a computer/robot would be capable of run-time object recognition and choosing appropriate activities.

V. Practical implementation

The Distributed Cognitive System is currently being implemented in a form of a distributed system managing a simple Semantic Network. There are two implementations being developed:

- OpenMP implementation on a single (but multiprocessor) workstation,
- Lam-MPI implementation for a computer cluster distributed environment.

The OpenMP-based implementation seems to be more convenient due to the ease of implementing and testing new algorithms (or other new ideas), without the necessity of using the time and resources of our computer cluster. It is easy to change topologies, to add/remove information input hardware, to "transmit" and process information. These issues become more complicated/ /challenging when developing a distributed application for a computer cluster.

On the other hand, running the developed algorithms on a low-cost heterogeneous computer cluster could enable programmers to spot any scalability issues and to improve/advance the ideas (theory and test implementations) to find a better solution. The Lam-MPI –based implementation becomes even more important than the OpenMP-based one, because it makes possible to design a truly scalable system. The scalability of the system becomes the key issue when the application is run – the amount of data to be transferred and stored in the knowledge base of the system usually becomes a severe test for the developed system [7, 8].

VI. Biologically-inspired Vision System

Human brain consists of relatively slow computational resources – neurons can not compete with computers when it comes to speed. However, researchers are still not successful in development of "artificial brain". Nevertheless, neurons are still much more efficient than computers. Their computational efficiency relays strongly on the high degree of parallelization of hierarchical networks [11]. The learning efficiency of neurons is also linked with parallelization of processing and their hierarchy – learning new data (tasks/objects/etc.) can be achieved using some already known sub-tasks/features – the re-usability of higher abstract representations (/features) is the key issue for speeding up the learning process.

Processing visual information in human brain, in this context, is similar to processing any other information – it also takes advantage of hierarchical networks parallelization. According to [11], *mainstream computer vision seems to follow design principles that are quite different from* its biological origin and *this general trend is worth reconsidering*, to *develop flexible and multipurpose vision modules that can contribute to a hierarchical architecture for artificial vision system*.

The Vision-based D.C.S., briefly described in this paper, contributes to this vision by showing the path to processing visual data in a cognitive system in a computer cluster.

VII. Conclusions

In this paper authors have briefly shown their innovative approach on the subject of computer vision systems dedicated for human-centered vision-based HCI. The novelty of the proposed solution relies on using vision as the main information input in various robotic and computer systems. In this case an innovative and efficient image processing was applied.

This paper shows the significance of the potential implementation of biologically reasoned computer vision in inter alia mobile robots, as vision is the most reliable information source for humans [3, 9, 10].

The interest in this research area was awaken by the need of improvement of already existing, similar solution in order to enable better 'imitation' of nature and to enhance (as mentioned above) other solutions, so the artificial intelligence would be as close to the 'natural' one as possible [7].

It is also important to mention wide potential application of the proposed Vision-based D.C.S – this is described in more detail in the following section.

VIII. Future Work

Human-centered Vision-based Human Computer Interaction Systems are becoming more visible in scientific literature and in everyday research. Robots are able to recognize movement in the scene, to follow an object, or to detect faces in the scene, but these tasks are just simple image processing algorithms. There are some mobile robots software implementations capable of learning and recognizing new objects (see [1]), but other – more challenging – algorithms (e.g. recognition of unknown objects and their autonomous classification) require more computational power and/or time and can not be implemented (embedded) on-board efficiently. Therefore, a Distributed Cognitive System can be applied to extend the cognitive abilities of an HCI-oriented vision-based robot.

Further research and implementations involve precise design and practical implementation of the Distributed Cognitive System.

There are also further plans to combine the computer vision with advanced signal processing for the purpose of design and development of a prosthetic arm. This is possible as the robots occupy the same physical space as human beings and therefore they are able to manipulate some of the same objects as we do. The most recent advances in robotics have enhanced their autonomy and as a result – their ability to work in non-deterministic environments and perform various task while not being instructed to do so, and/or not being taught precisely how to do it.

Constantly increasing progress in research enables to combine the bio-signals with CV in order to create and/or improve bio-navigation in many life areas [10].

REFERENCES

[1] Körner E., "Elements of Brain-like intelligence for ASIMO - Learning visually guided autonomous interaction". Japan Society for the Promotion of Science, online, accessed 2013-10: http://www.jsps-bonn.de/fileadmin/veranstaltungen/2009_Abstr_Koerner.pdf

[2] Podpora M., Sadecki J., "Biologically reasoned point-of-interest image compression for mobile robots", in: Hippe Z., Kulikowski J. (eds.), "Human-Computer Systems Interaction – Backgrounds and Applications", chapter IV – "Robots and Training Systems", ISBN: 978-3-642-03201-1, DOI: 10.1007/978-3-642-03202-8_29, Springer, pp.389-401, 2009

[3] Podpora M., "Vision processing for autonomous robots with the use of distributed systems" (in Polish), [doctoral thesis], Opole University of Technology, 2012, available online, accessed 2013-10: http://www.dbc.wroc.pl/dlibra/docmetadata?id=13866

[4] Michalski R. S., "Knowledge mining: a proposed new direction", Sanken Symposium on Data Mining and Semantic Web, Osaka University, Japan, March 10-11, 2003

[5] Fayyad U., Piatetsky-Shapiro G., Smyth P., "From data mining to knowledge discovery in databases", AI Magazine, pp.37-54, 1996

[6] Frawley W., Piatetsky-Shapiro G., Matheus C., "Knowledge discovery in databases: an overview", AI Magazine, Vol 13, No. 3, pp.57-70, 1992

[7] Podpora M., Kawala-Janik A., Pelc M., "Policy-based Self-configuration of Autonomous Systems Information Inputs", Proceedings of the 2013 IEEE 7th Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), vol.2, ISBN: 978-1-4799-1426-5, Berlin 2013

[8] Ward P., Pelc M., Hawthorne J., Anthony R., "Embedding dynamic behaviour into a self-configuring software system", ATC 2008, Lecture Notes in Computer Science, Springer Verlag, 2008

[9] Moehrmann J., Heidemann G., "Efficeint Annotation of Image Data Sets for Computer Vision Applications", VIGTA'12, In Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications, ACM, New York, 2012

[10] Lavely E., Meltzner G., Thompson R., "Integrating Human and Computer Vision with EEG Toward the Control of a Prosthetic Arm", HRI'12, In Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, ACM, New York, 2012

[11] Kruger N., Janssen P., Kalgan S., Lappe M., Leonardis A., Piater J., Rodriguez-Sanchez A.J., Wiskott L.; "Deep Hierarchies in the Primate Visual Cortex: What Can We Learn for Computer Vision?", In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, Issue 8, IEEE 2013