Dynamic re-definition of Region-of-Interest in Vision System's Feedback

Michał A. Podpora, *Student Member, IEEE*

***Abstract*—In this paper the author describes a modification of the Region-of-Interest idea, in which the "best quality" area is defined individually for every frame of video stream. The robot's vision system decides which fragment of the whole image should be transmitted using the best quality.**

***Index Terms*—active vision, perceptual reasoning, representations and transforms, scene analysis**

I. Introduction

THE input video stream of a mobile robot's vision system is known to carry too much information to be processed by a computer cluster. Although it is possible to use data compression, the traffic among nodes of a cluster is still far too heavy. Lossless compression is not efficient enough and lossy compression is not acceptable for any image understanding algorithms. One solution to this problem is "active vision", in which only a fragment of the whole image is processed. This is fair enough for recognizing license plates' numbers but it is too simplified to be used in an autonomous robot. A mobile robot should be able to observe more than analyzed detail of an object.

In this paper the author considers using the Region-of-Interest (ROI) idea, described in [1], for every frame of mobile robot's video stream. An image, inside the so-called ROI area, contains data compressed using a better image quality. Alternatively, the ROI area can be compressed using lossless compression. However, it complicates all the image compression, decompression and processing algorithms.

II. Region of Interest

Originally the ROI was designed to define a rectangular area of better image quality. Usually it is used for better feature representation in selected regions. ROI enables the ability to compress an image with average or poor quality and still have the most important fragment of image compressed with good quality [2]. Despite the data size reduction, the image is still useful for pattern recognition or inference.

Work co-financed by European Social Fund.

M. A. Podpora is with the Department of Electrical and Computer Engineering, Opole University of Technology, ul.Sosnkowskiego 31, 45-272 Opole, Poland, phone: +48 604 426744; fax: +48 77 4006338; e-mail: michal.podpora@gmail.com.

III. Active Vision

Active Vision (AV) is often understood as analysis of a part of an image while ignoring the rest of it. Alternatively, some implementations use the AV idea for scanning and processing the whole image tile-by-tile [3]. In the latter case, the rest of the analyzed image, in a given moment in time, is also being ignored.

Although initially AV was based on human eye, all the resemblance was limited to data reducing similarities. Analysis of a whole image was either complicated because only a fragment of image was acquired or because an image, or a part of it, was represented by a set of acquired fragments.

The combination of Region-of-Interest and Active Vision gives the ability to process a high quality fragment of image while not loosing the rest of information. In [4] the author describes an Active Vision implementation, using spatially variable quality, depending on the distance from currently processed detail. In his project neither ROI conception nor JPEG2000 compression was used.

The most important advantage of using the junction of ROI and AV is the ability to analyze a detail (fragment) of an object and, at the same time, to notice that another detail exists in the image (Fig. 1). Unlike in AV, the vision system does not ignore the rest of the image. Instead, the outside of ROI is being partially used for choosing new ROI position and size.

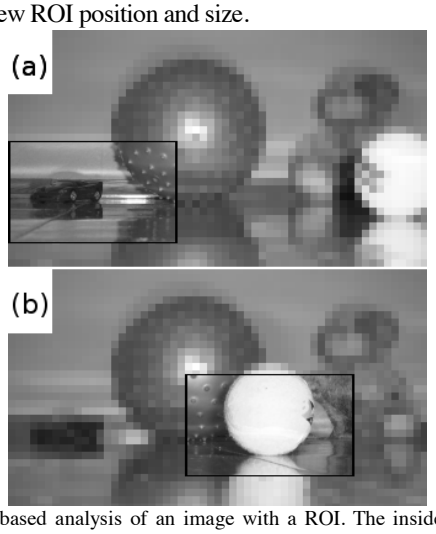

Fig. 1. AV-based analysis of an image with a ROI. The inside of the ROI rectangle is being processed, while the outside of ROI is compressed with quality good enough to notice new objects. The vision system was examining the red car (a), when the tennis ball has entered the scene (b). While the ball is a relatively new object, it attracts the system's attention.

978-1-4244-3898-3/09/$25.00 ©2009 IEEE

IV. Feedback

In the practical implementation, new location (coordinates) of the fixed-size ROI is defined separately for every frame of the video input stream. The election of next frame's ROI is performed on the basis of few criteria/ratings, like movement, color or contrast.

In future version of the software, with some basic Artificial Intelligence (AI) algorithms included, a new mighty criterion will be implemented: the decision of the AI subsystem (Fig. 2).

The elections are essential for the whole conception, while they enable the ability to use the feedback to change parameters (i.e. ROI) of the next frame.

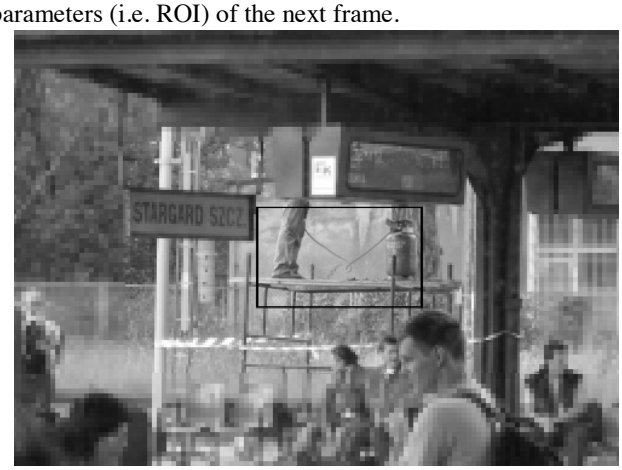

Fig. 2. The ROI location election algorithm should be not limited to contrasting or moving objects [Future work]. The "decision" of the AI algorithm should be primary. For a human, the most important part of this picture in most cases is the safety of the worker. For a today's vision system it is impossible to "focus" on this part of the image because of e.g. movement of other people in the scene.

V. HTM

The developed application is intended to use the Hierarchical Temporal Memory (HTM) in future releases. The idea of HTM network's understanding of the input data/signal is based upon human brain's neocortex [5].

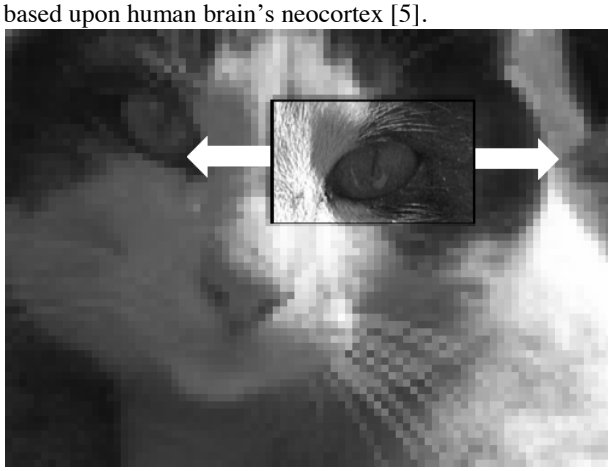

Fig. 3. Step-by-step high-level pattern spatial coexistence verification. The vision system recognizes an eye and then the algorithm "decides" to check if there is another eye in the image, to prove that the object should be recognized as a face. A face may be a part of a human or an animal; it may contain some characteristic features.

HTM input nodes acquire data and build higher abstraction levels in subsequent layers. Finally, a "belief" is formed, which can be represented, for instance, as an object. HTMs are described in more detail in [5]–[7].

Human do not acquire whole images like a camera. In fact, every object representation is an outcome of sequentially acquired details at various fixation points.

Because neocortex was the inspiration for HTM inventors, usage of the proposed vision system is doubtlessly convenient.

VI. Conclusion

Using the AV ideology by analyzing dynamically redefined ROIs seems to be the natural way of perceiving. Creating a robot capable of human-like human-machine interaction is not possible if the robot does not "understand" our world "our way", that is, by creating spatio-temporal memory hierarchies and instances, which we simply call memories.

Most important advantages of using the proposed idea are:

- Analysis of only a part of an image, which is a typical AV approach, significantly speeds up computations.
- The image quality is superior due to usage of ROI.
- The image is compressed to give smaller data size [8].
- It is possible to spot objects/details outside ROI.
- Vision system has full control over the location of ROI.

Acknowledgment

The author thanks Jan Sadecki and Marian Łukaniszyn for valuable suggestions, support and encouragement in the research progress.

References

[1] ISO / IEC FCD 15444-1 : 2000, "JPEG2000 image coding system", Annex H, Available online: http://www.jpeg.org/public/fcd15444-1.pdf Accessed 30 March 2009.

[2] D. Taubman, M. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, 2004, pp. 415–416 and 442–449.

[3] M. Nałęcz, W. Duch, J. Korbicz, L. Rutkowski, R. Tadeusiewicz, *Biocybernetics and biomedical engineering 2000 – vol 6 – neural networks*. Akademicka Oficyna Wydawnicza EXIT, Warsaw, 2000, ISBN 83-87674-18-4, (in Polish), pp. 482–493.

[4] M. Podpora, „Can the decrease of the amount of humanoid vision system's data increase its capabilities?", in: *Control and Automation: Present Problems and Solutions, chapter XI – Artificial Intelligence*. Oficyna Wydawnicza EXIT, Warsaw, 2008, ISBN 978-83-60434-42-0, (in Polish), pp. 680–687.

[5] J. Hawkins, S. Blakeslee, *On Intelligence*, Times Books, 2004, ISBN 080507456-2, ISBN 978-080507456-7.

[6] M. Podpora, „HTM – a new age in Artificial Neural Networks history?", Oficyna Wydawnicza Politechniki Opolskiej, (in Polish), submitted for publication.

[7] J. Hawkins, G. Dileep, *Hierarchical Temporal Memory - Concepts, Theory, and Terminology*, Numenta Inc., 2007, Available online: http://www.numenta.com/Numenta_HTM_Concepts.pdf Accessed 31 March 2009.

[8] Podpora M., "Biologically reasoned machine vision: RLE vs. entropy-coding compression of DWT-transformed images", *Proceedings of the 14th Conference Student EEICT 2008 vol.4*, Brno, 2008, ISBN 978-80-214-3617-6, pp. 457-460.