BIO-INSPIRED FLIGHT CONTROL AND VISUAL SEARCH WITH CNN TECHNOLOGY

Csaba Rekeczky, Dávid Bálya, Gergely Timár, and István Szatmári

AnaLogical and Neural Computing Systems Laboratory
Computer and Automation Institute of the Hungarian Academy of Sciences
1111 Budapest, Kende u. 13-17, Hungary

ABSTRACT

In this paper it is shown that by building on parallel topographic CNN preprocessing of image flows, efficient terrain exploration and visual navigation algorithms can be developed. The approach combines several channels of nonlinear spatio-temporal feature detectors within an analogic CNN algorithm and produces unique binary maps of salient feature locations. This preprocessing scheme is embedded into a multi-target tracking (MTT) framework where these features are statistically described and assigned to numbered tracks. The MTT output has two distinct roles. First, its feature descriptors drive a classifier based on the adaptive resonance theory (ART), which is also implemented on CNN architecture. Second, it provides an optical flow ("target displacement") estimate to the navigation system, which in turn calculates the flight control parameters (Yaw-Pitch-Roll). An upper level visual attention and selection mechanism uses both the feature descriptors and the optical flow estimates to automatically adjust the focus and scale (zoom) during navigation. The paper describes the architecture and the algorithmic frameworks and provides the first experimental results on aerial video-flows.

1. INTRODUCTION

The intention of bio-inspired engineering of exploration systems is to learn the principles found in successful, nature-tested mechanisms of specific "crucial functions" that are difficult to accomplish by conventional methods, but which are realized rather effectively in nature by biological organisms. The intent is not just to mimic operational mechanisms found in specific species but also to learn the salient principles from a variety of diverse bio-organisms for a desired "crucial function". We are deciphering many of these natural visual strategies and have found ways to apply the results in areas such as navigation, stable flight and terrain following. With the use of biomorphic fliers [1] our results could also contribute to previously impossible projects in related fields of science.

Recent biological studies have confirmed that representations of each different characteristic of the visual world are formed in parallel, and embodied in a stack of "strata" in the retina [8]. Each of these representations can be efficiently modeled in Cellular Nonlinear Networks (CNN, [2]-[4]). When translated into CNN image processing operations, many of the biological functions constitute algorithmic cornerstones, useful in practical applications. It is our view that incorporating the success strategies of bio-inspired navigation and visual search/pattern recognition/image understanding into engineering solutions should lead to rapid advances in future missions. Our work primarily focuses on architecture and algorithmic framework design and makes the first steps toward a CNN based system-on-a-chip (SoC) for visual search and navigation tasks.

During this work we have been developing stored program cellular nonlinear processing strategies for terrain exploration and classification; automatic adjustment of focus and scale of attention; navigation support and a system level capability to track multiple targets. The experiments have been conducted mainly on monocular aerial video-flows showing diverse terrain sites from a navigating plane.

2. MULTI-TASK SYSTEM DESIGN

2.1 System Description

The general system architecture designed for bio-inspired visual search and navigation is shown in Fig. 1. Assuming a large resolution array sensor input the focus and scale is automatically adjusted in a feedback loop depending on feature processing results. The selected sensor input undergoes a parallel multi-channel CNN processing, which provides a topographic (binary) output for the multi-target tracking (MTT, [7]) framework. The output of the MTT sub-system consists of static and dynamic target attributes. In this context, targets could be both salient terrain features (e.g. river forks, irregular/large rocks etc.) and objects in motion (e.g. flying air-vehicles, birds etc.). Static target attributes include target feature descriptors such as centroid locations, contour and skeleton structure, orientation, size and others. Dynamic target attributes describe all targets in motion with their kinematic properties (this includes the optical flow: the estimate of the 2D motion field). The MTT core is driven by the local feature extraction module. Based on the local feature descriptors at a specific frame, a distance is calculated to all existing targets. Then, filtered by a tunable gating mechanism these distances in the feature space are taken as the likelihood of whether an actual object belongs to a specific target track. The arrangement decision is made in the data association module, which is also responsible for adjusting the main algorithm parameters of the analogic CNN preprocessing. The output of this module is the input to the state estimation block where based on kinematic assumptions (refined through subsequent measurements) the target states are predicted for the next frame and used as a reference for the distance calculation block.

This system description envisions a platform with multi-task processing capability. Such architecture have been designed and built; and is referred to as the Compact Cellular Visual Microprocessor (COMPACT CVM). The COMPACT CVM architecture (Fig. 2) builds on state-of-the-art CVM type (ACE16k) and DSP type (TEXAS 6x) microprocessors and its

0-7803-7762-1/03/$17.00 ©2003 IEEE

algorithmic framework contains several feedback and automatic control mechanisms in between different processing stages (Fig. 3). The architecture is standalone and with the interfacing communication processor it is capable of 100 Mbit/sec information exchange with the environment (over TCP/IP). The COMPACT CVM is also reconfigurable, i.e. it can be used as a monocular or a binocular device with a proper selection of a large resolution CMOS sensor (IBIS 5) and a low resolution CNN/CVM sensor-processor (ACE16k).

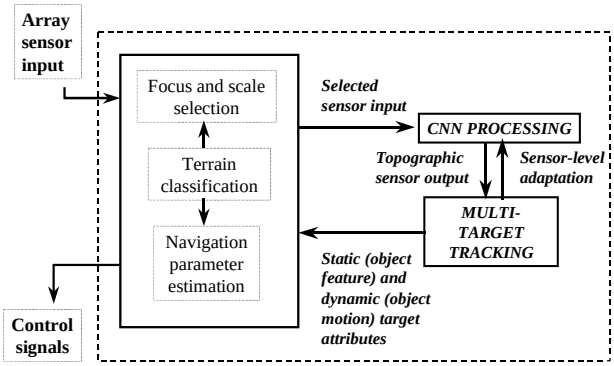

Fig. 1. Main processing blocks and signal flow of the visual microprocessor architecture designed for bio-inspired visual search/navigation

The COMPACT CVM architecture is a fault tolerant multi-task visual computer for bio-inspired exploration, selection, tracking and navigation. We believe that this architecture also provides a framework for a SoC design: toward a fully integrated cellular sensor-computer.

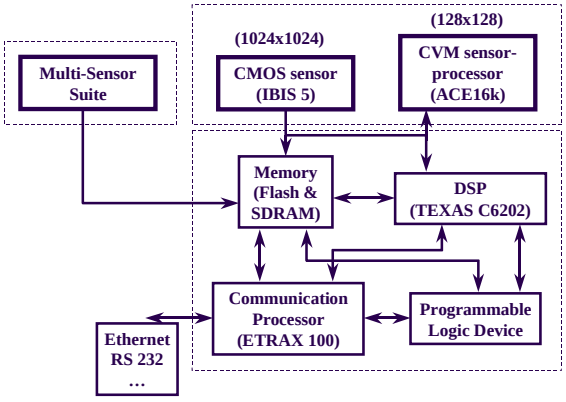

Fig. 2. Main building blocks of the COMPACT CVM architecture

2.2 Biological Motivations

There is a strong biological motivation behind building a multi-channel adaptive algorithmic framework for visual search and navigation. It has been long known that the mammalian visual system processes the world through a set of separate spatio-temporal channels. A recent study has confirmed that the organization of these channels begins in the retina where a vertical interaction across ten parallel stack representations can also be identified ([8]).

Beyond reflecting the biological motivations our main goal was to create an efficient algorithmic framework for real-life experiments, thus we have decomposed the above model into a multi-channel adaptive (single-layer) CNN-UM analogic algorithm. Within this algorithmic framework the enhanced image flow is analyzed by temporal, spatial and spatio-temporal filters. The output of these sub-channels are then combined in a programmable configuration to form the new channel responses. Crisp or fuzzy logic strategies define the global channel interaction and result in a unique binary image flow. This is also combined with the output of a single/multiple step prediction and forms the final output.

When building up the computing blocks of the above multi-channel algorithmic framework the following key processing strategies learned from retina modeling and biological vision related experiments have been used (the associated image processing / system design arguments are given in italic):

- spatial, temporal and spatio-temporal decomposition of the input flow: *an efficient geometric distortion analysis requires a sparse signal representation*
- signal flow normalization: *dynamic range optimization*
- parallel on-off channel processing: *DC-component compensation*
- narrow and wide-field wave-type interaction: *efficient binary patch shaping with noise suppression*
- "vertical" interaction of the decomposed channels: *forming a unique detection output through optimized "cross-talk" of the individual channels*
- attention and selection mechanisms*: efficient content and context dependent processing*
- saccade detection mechanisms*: proper handling of large shifts in the field of view*

3. THE ALGORITHMIC FRAMEWORK

In the COMPACT CVM systems there is an upper-level visual attention and selection mechanism adjusting the focus and scale (zoom) of processing ([10]). This mechanism can select a single or multiple windows from the same large resolution frame at a certain time instant, however the video-flow is never processed in parallel at full resolution (much alike biological systems). This framework allows the system of multi-task execution at different time scales. The general algorithmic framework (Fig. 3) incorporates optical flow estimation, feature classification and automatic attention mechanisms built on topographic nonlinear parallel feature processing performed by the CNN/CVM sensor-processor.

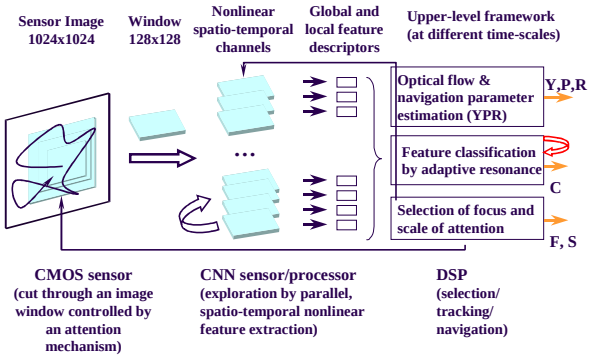

Fig. 3. Flow-processing diagram of the COMPACT CVM algorithmic framework

Topographic Multi-Channel Preprocessing

In Fig. 4 examples are shown for calculating the topographic feature maps (the first parallel stage of the analogic algorithm described in Fig. 3) for terrain video-flows. As illustrated (see the right most column of the figure) these maps describe the edginess, irregularity, rough/fine structures, connected structures etc. of the input. It is also obvious that the description is non-orthonormal in the feature space thus providing certain robustness when some of these specific estimation strategies are sensitive and not reliable enough for further processing.

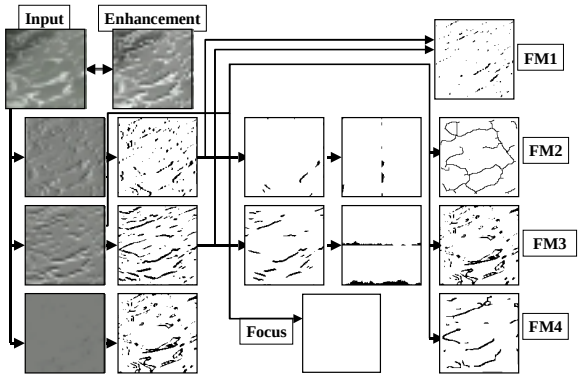

Fig. 4. Example for topographic feature maps (FM) calculated by CNN processing.

3.1 Attention and Selection Mechanisms

If the sensory input is a large resolution array video-flow and only a certain region of interest is evaluated, an automatic mechanism is needed to select the appropriate focus and scale of processing (Fig. 5). In the current experiments we are using video-flows acquired by large resolution sensors and implementing various attention and selection mechanisms (see e.g. [10]) based primarily on local target descriptors and secondarily on motion-flow estimates. As it has been noted in the earlier sections, the configuration of such a mechanism is dependent on a classifier output ("different strategies for radically different terrains").

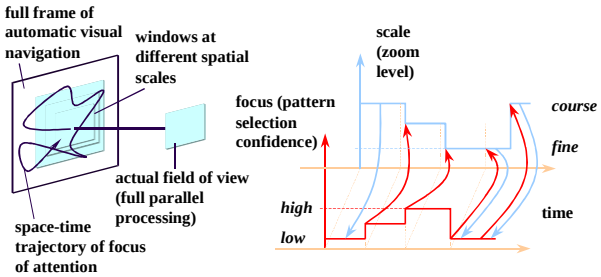

Fig. 5. Driving the focus and scale of attention with a mutual interaction of the focus and scale selection mechanisms.

3.2 Feature Based Classification Schemes

There are several classifiers that could have been used for terrain feature analysis and global classification. We have applied an adaptive resonance theory (ART, [13]) based module capable of learning on a pre-selected image flows (training set) and performing the classification in a dynamic procedure (new classes will also be formed "on-line" if the test image flows are not similar to the learnt categories). The process description of off-line supervised learning and on-line classification is shown in Fig. 6. The ART network has its roots in neurobiological modeling and has a strong mathematical background (e.g. modeling THALAMUS V1). It is a further advantage that a modified version of ART can be implemented on existing CNN-UM architecture. Thus, the combined CNN-ART algorithmic framework is a fully biologically inspired algorithmic approach to terrain feature analysis implemented in our current system.

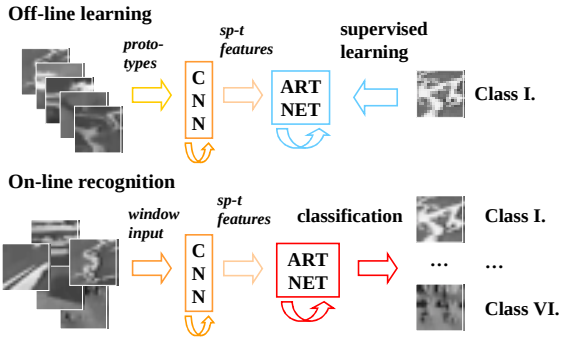

Fig. 6. Terrain exploration and classification with off-line supervised learning and on-line recognition. The topographic processing (spatio-temporal feature extraction) is completed by a CNN block, while the feature vector learning and classification is performed by an ART network.

3.3 Navigation Parameter Estimation

An integrated motion field and flight control parameter estimation algorithmic framework is shown in Fig. 7. As described in the introductory section the solution for visual navigation is also based on a multi-target tracking layer driven by multi-channel CNN preprocessing. In its simplified form (see flow description I. in Fig. 7) this could be a multi-feature tracking over a fixed uniform grid (no dynamic descriptors) or the enhanced version of the multi-target tracking scheme over an adaptive non-uniform grid (see flow description II. in Fig. 7). The output of both types of processing should undergo a robust nonlinear sorting, translation and scaling before it enters the stage of navigation parameter estimation.

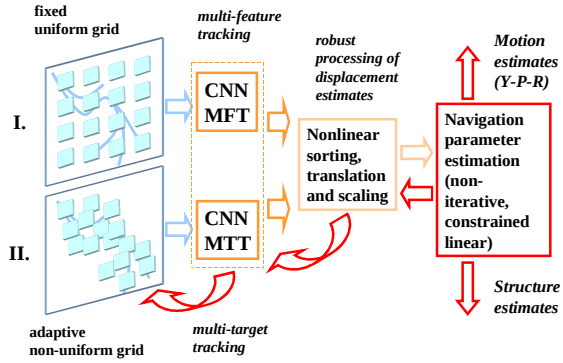

Fig. 7. Motion field and flight control parameter estimation based on a fixed uniform and/or adaptive non-uniform grid defined over the large resolution video-flow.

The flight control parameter estimation module requires the motion field estimates (the optical flow) for each frame as an input and it calculates the unit translational direction along with the rotation parameter estimates (and possibly the structure estimates). We have found that reliable motion field estimates can be derived from the MTT output if a reasonable number of features are tracked in parallel (typically more than a dozen, see also [11], [12]).

4. EXPERIMENTAL RESULTS

While we have started the measurements of the COMPACT CVM with the ACE16k [6], the architecture and the algorithmic framework described in this paper have been emulated within the ACE-BOX environment [9] hosting the previous generation CNN-UM type microprocessor (ACE4k, [5]). This system is capable of processing at a video frame-rate. Samples from the first experimental results are shown in Fig. 8 and Fig. 9.

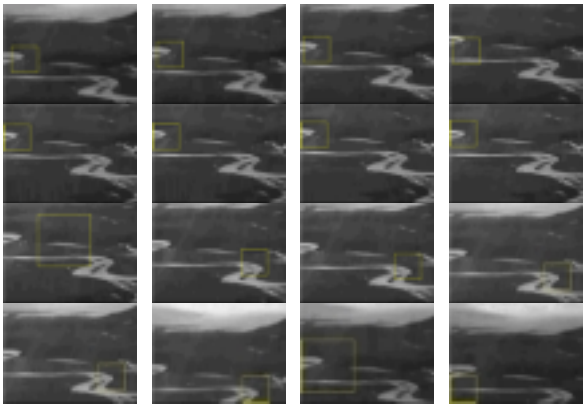

Fig. 8. This image sequence shows the operation of the attention-selection mechanism driven by the classifier output. The algorithm attempts to focus on the most river-like area of the input video flow. Observe, that the selected view port is enlarged periodically to detect relevant locations that lie outside the area in focus. The white square shows the area of the input video that is in focus, i.e. scaled down to the resolution of the visual microprocessor.

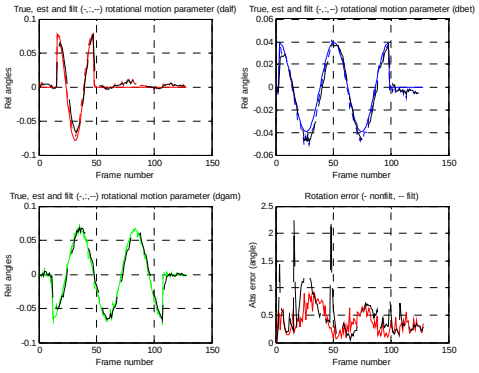

Fig. 9. Experimental results of navigation parameter (Y-P-R) estimation compared to ground-truth references.

5. SUMMARY

We have proposed a CNN technology based bio-inspired visual microprocessor architecture and a dedicated algorithmic framework for efficient terrain exploration, site selection, tracking and navigation.

This work was partially funded by the ONR (TeraOps Ltd CA USA, No: 0295-R, sub-contract) and by the Hungarian National Research and Development Program, Grant No. NKFP 035/02/2001.

6. REFERENCES

[1] S. Thakoor, J. Chahl, M. V. Srinivasan, L. Young, F. Werblin, B. Hine, and S. Zornetzer, "Bioinspired Engineering of Exploration Systems for NASA and DoD", invited talk, presented at Artificial Life VIII., 2002.

[2] L. O. Chua, “CNN: a Vision of Complexity ”, Int. J. of Bifurcation & Chaos, Vol. 7, No. 10, pp. 2219-2425, 1997.

[3] T. Roska and L. O. Chua, "The CNN Universal Machine", IEEE Trans. on Circ. & Syst., Vol. 40, pp. 163-173, 1993.

[4] F. S. Werblin, T. Roska, and L. O. Chua, "The Analogic CNN Universal Machine as a Bionic Eye", International Journal of Circuit Theory and Applications, Vol. 23, pp. 541-569, 1995.

[5] S. Espejo, R. Domínguez-Castro, G. Liñán, Á. Rodríguez-Vázquez, “A 64×64 CNN Universal Chip with Analog and Digital I/O”, in Proc. ICECS’98, pp. 203-206, Lisbon 1998.

[6] G. Liñán, R. Domínguez-Castro, S. Espejo, A. Rodríguez-Vázquez, “ACE16k: A Programmable Focal Plane Vision Processor with 128 x 128 Resolution”, ECCTD ’01, pp.: 345-348, August, Espoo, 2001.

[7] S. Blackman and R. Popoli, "Design and Analysis of Modern Tracking Systems", Artech House, 1999.

[8] B. Roska and F. S. Werblin, “Vertical Interactions across Ten Parallel Stacked Representations in Mammalian Retina”, Nature, 410 pp 583-587, 2001.

[9] AnaLogic Computers Ltd: Aladdin Pro R2.3, http://www.analogic-computers.com/, Budapest 2002.

[10] E. Niebur and C. Koch, "Computational Architectures for Attention", R. Parasuraman (ed.) The Attentive Brain, Cambridge, MA:MIT Press, pp. 163-186, 1998.

[11] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, “Performance of Optical Flow Techniques”, International Journal of Computer Vision, Vol. 12, No.1, pp 43-77, 1994.

[12] R. I. Hartley, “In Defence of the 8-point Algorithm”, 5th International Conference on Computer Vision (ICCV’95), pp. 586-593, 1998.

[13] G. Carpanter and S. Grossberg, “A Massively Parallel Architecture for a Self-organizing Neural Pattern Recognition Machine”, Computer Vision, Graphics, and Image Processing, Vol. 37, pp. 54-115, 1987.