

GVPP: A Generic Technology with Sense & Avoid Applications. 4th. European micro-UAV Meeting 15-17 September 2004 Toulouse France

Patrick Pirim

BEV Development, Luxembourg

E-mail: patrick.pirim@bev.lu

Abstract: We postulate that humans, through high-level orders that they will autonomously and adaptively execute, will command the UAV of the future. We want to endow these UAV with some capacities for autonomy and adaptation that characterize mere animals in general and birds in particular. Such capacities will call upon both integrated multimodal sensory perception – e.g. to orient to important stimuli - and cognitive skills - e.g., to learn and memorize a given situation or event, or to plan a trajectory, with obstacles avoidance by optical flow:

The UAV tends to stay in the middle of narrow corridors, while its forward velocity is automatically reduced when the obstacle density increases. Moreover, when heading into a frontal obstacle, the understanding is able to generate a tight U-turn that ensures the UAV's survival.

Key words: optic flow, obstacle-avoidance, GVPP,

1. Introduction

In its current stage of development, the biomimetic design of GVPP chip produced by BEV affords it the unique feature of exhibiting four of the main adaptive capacities that characterize natural vision.

[1] Real-time information processing: through circuits dedicated to movement, speed, colour, hue and luminance, oriented edges, corners, the system monitors visual inputs on-line, thus

making quick reactions and efficient predictions possible. Successive images are not recorded but processed on-line according to previous-step anticipations.

[2] Detection of objects or events through temporal coincidences and unsupervised collective decisions: to detect a landmark or an object within an ever-changing flow of images entails being able of singularising sets of coherent data within such flow. Interesting structures within an image are characterized by temporal coincidences within different circuits (e.g., the simultaneous occurrence of colored and moving zones). Through successive images, different modules of the chip take collective decisions about these structures, notably about their positions or movements.

[3] Object pursuit and anticipation: the detection of an object within an image is difficult because of permanent changes caused by displacements or light variations. The activation of the different modules in the system is permanently adapted to changes in the appearance of an object. Moreover, these modules cooperate to anticipate the content of future images and to track in parallel several moving objects - thanks to dedicated circuits, like those that detect movements.

[4] Learning of visual characteristics: adaptive and real-time tracking of coherent information sets makes it possible to characterize and record the intrinsic properties of given objects. Later on, these objects may be retrieved or recognized, even in case of temporary disappearance from one or several successive images.

2. Why GVPP?

To perform a vision application, it is necessary to use **a generic technology to perceive, understand and act in real time.**

Instead of the understanding power concentration of Imaging Processing (IP) we prefer to focalise on the computation on perception; the more you perceive more the understanding is easy.

Even today, we solve the application with a small Generic Visual Perception Processor when the IP needs a strong Digital Signal Processor.

The global GVPP size, power, and time developments are effectively lower with a better result



Fig. 1. GVPP7-V Analog video Input.



Fig. 2. GVPP7-C Cmos Imager Included

3. GVPP model

From Buser and Imbert [2] physiological description, we have developed an electronic concept using:

- Environment descriptions by spatio-temporal perception, which can be translated in sequences, the sample rate determine the dynamic, and each sequence is sampled in sub-sequences as rows and columns in visual perception. The sequences describe the temporal information and the sub-sequences the spatial information. Two main electronic blocks: Temporal Domain Computation associated to Spatial Domain Computation extract features.
- A duality between a low physiological neuron speed function including a huge synaptic connections and a high speed silicon electronic computer with only one multiplexing connection.

These electronic circuits are directly inspired by the properties of the human visual system including the retina, the low and high-level visual cortical areas. The basic principle of these circuits is to cumulate within the same chip the set of adaptive properties existing in the human visual system. This includes adapting the sensitivity to the background light, tracking relative movements of objects or people, anticipating their movements, and learning to detect objects or people, whatever their position, size, orientation, and whatever their backgrounds. The processing principles in such electronic circuits are inspired from processing principles observed in populations of retinal and cortical neurons. They do not aim at a complete neuronal implementation, which is very costly in electronic circuits, but rather capture simply as possible the adaptive properties of the neural processes.

The Spatio-Temporal Neuron bloc modelling by electronic implementation schematic as describe bellow, have:

1. A sample rate frame (sequence) and pixel (sub-sequence) for temporal and spatial cues perception.
2. Each frame or sequence has an automat for, initialisation, flow computation and results analysis.
3. A functional group of 2 STNs (Spatio-temporal Neuron) receives one temporal and one spatial cue.

- Each STN is implemented with a histogram computation, a classification unit and a time-coincidence unit. The STN is a duality of a physiological neurons population with the same properties as population analysis, emergence of a majority vote, amplification by time-coincidence, and prediction.

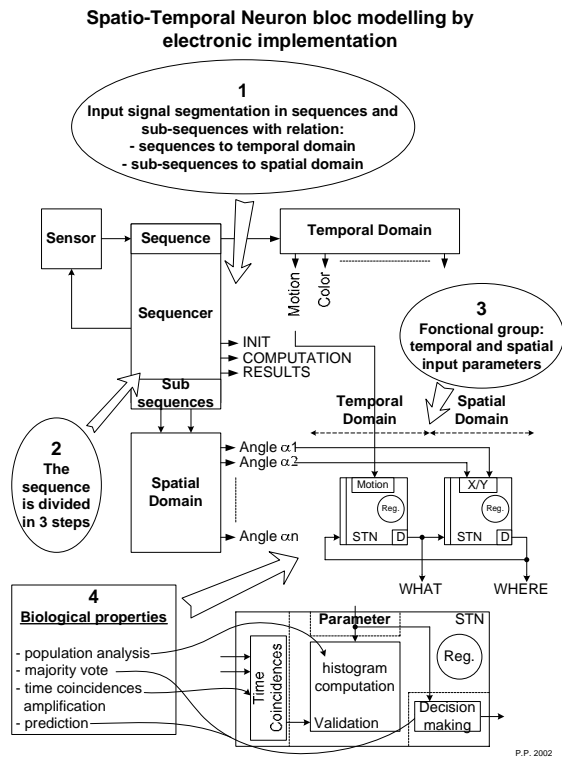


Fig. 3. GVPP model

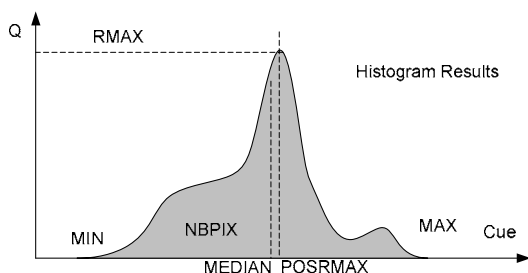
In each end of frame, the histogram computation upgrades STN registers:

NBPIX is the number of values comprise in the histogram computation. It is also the energy of the function.

RMAX is maximal quantity

POSRMAX is the cue's value of this maximal quantity.

MEDIAN is the cue's value of the middle of the histogram.



4. Program

- At the initialisation, a specific connection between a minimum of two STNs is done by writing registers on STN with a temporal and spatial cues.

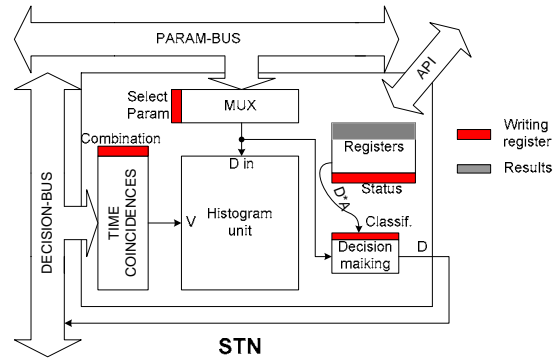


Fig. 4. The generic block: STN for Spatio Temporal Neuron

For example, to perceive motion on a screen, on the first STN a temporal cue (motion) is selected as input parameter in all range on the classification unit. The time-coincidence receives the classification result of the second STN, which receives a spatial cue as input parameter in the entire screen on the classification unit. The time-coincidence receives the classification result of the first STN.

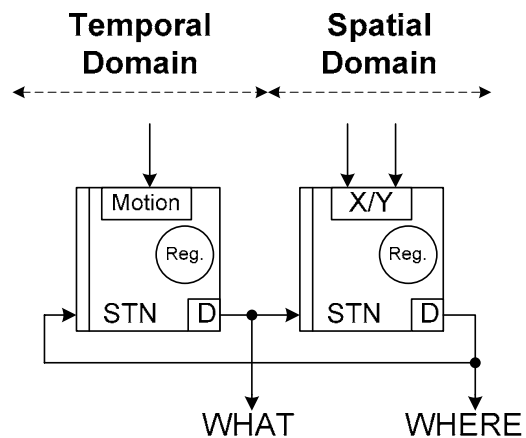


Fig. 5. "What and Where" processing, as tracing cortical connections suggests that the primate auditory system, like the visual and somatosensory systems, may be organized into "what and where" pathways. Jon H. Kass and Troy A. Hackett. Nat. Neurosci. 2, 1045-1046 (1999).

- On end of recurrent frames, a test is done on NBPIX register, or the value is under a threshold (noise), and all the classification values are initialised, otherwise an automatic classification is done and the motion is tracked. The **MEDIAN** registers are the barycenter of position (ROI region of interest) and level of motion. This task works till is killed.

5. Example of application

To perform applications, two others functions are necessary:

Dynamic recruitment, the program described before does one task, the dynamic recruitment of other STN group [14] allows perceiving environment.

Inhibition, it is a very important function, by inversion signal on a same scale; it permits object segmentation.

For example, to perceive a face, first motion is activated and when the energy is sufficient a recruitment of colour analysis permits to perceive the main colour of this first ROI. Inside this ROI the inhibition of other STN select the no-main colour. The barycenters of each ROI detected form a tree. This tree is a face representation [12, 13].

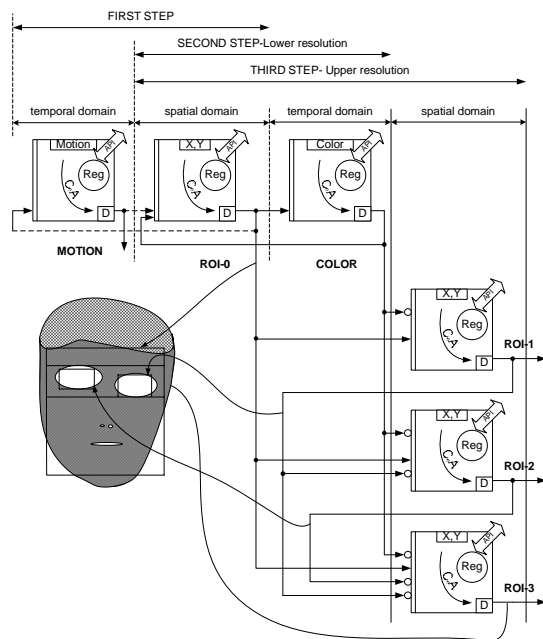


Fig. 6. Graphic application with STN

This *top down* representation encodes the object with invariance in translation and rotation. The fusion of accelerometer sensor allows the size invariance. The invariance's organization tree permits the *bottom up* research of object. It is a form of *re-entrance* predicted by Edelman.

6. Obstacle avoidance

What is optic flow?

- Global pattern of motion vectors for elements in the visual scene induced by moving a point of observation.
- Focus of Expansion (FOE) is the central point from which optic flow radiates.
- FOE specifies current heading.
- Detection of heading from optic flow is very accurate for translation

Two visual strategies for flying to a target

- Optic flow: To maintain FOE on target
- Egocentric direction: To fly in direction of target.

Observers may be exhibiting two different adaptive responses:

- Calibration motor: quick motor adjustment to visual information, no remapping of sensory-motor systems.
- Adaptive kinesthetics: gradual decrease in error as a result of repeated exposures, possible remapping of sensory-motor systems.

Optic flow dominates as it becomes more salient in the visual scene.

GVPP perceives, in real time, temporal variations for each pixel intensity to determine the threshold of detection sensibility. The spatial analysis of coherency noise determines the *motion field* from which information about the self-motion of the camera or about the structure of the scene can be inferred [11]. When such movements are sampled, by means of a video stream for instance, the apparent motion of pixels in the image constitutes the optic flow which is a convenient approximation of the motion field if the intensity of each pixel is preserved from one frame to the next.

In 1865, Von Helmholtz explained how flying insects are able to evaluate the distances of lateral objects by using motion parallax and, since that time, numerous studies have investigated the corresponding mechanisms. Srinivasan et al., [1, 17, 18] for instance, have studied how bees call upon a *centering*

response, which consists in equalizing the optic flow on left and right sides, to fly in the middle of a textured tunnel. Bees also exhibit a *clutter response* that enables their speed to be adapted to the width of the tunnel by maintaining a constant average motion. Likewise, Franceschini et al. [15, 16] described how the organization of the compound eye of the housefly, and how the neural processing of visual information obtained during the flight, allow this insect to compute its distances to lateral obstacles and to avoid them. It also appears that other scientists [4, 5, 6, 7] have understood how other animals use the frontal optic flow to estimate the so-called *time-to-contact*, i.e., the time before a frontal collision is likely to occur.

In the following, we use the UAV body-frame to refer to a point (x, y, z) in the environment and the classical image-frame to refer to a pixel (l, j) . The focus of expansion (FOE), which is the projection of the direction of the UAV onto the image, is taken as the center of the image frame. Parameter f refers to the focal length of the camera. We will use symbols M and V , respectively, to distinguish between motion on the image plane and velocity in the environment [9].

The component of the optic flow that is required to avoid obstacles is generated by a forward translation V_x of the observer. In this case, the horizontal motion M_h of an object is proportional to the inverse of its distance to the observer. From fig. 7, the following relationship may be derived:

$$h/Y = f/X \quad (1)$$

After differentiating with respect to time and substitutions, it finally appears that horizontal motion M_h of pixel is given by the equation:

$$M_h = h \cdot V_x / (d \cdot \cos\beta) \quad (2)$$

From Eq. (2), we can deduce a strategy equalizing the perceived pixel motions will tend to maintain equal the distances to obstacles on both sides of the UAV. This strategy was called either the *balance strategy* or the *centering response*.

Two STNs compute the average horizontal motion of pixels, respectively, on the right and on the left. The UAV must turn according to a value proportional to the difference of motions measured on both sides to reduce this difference.

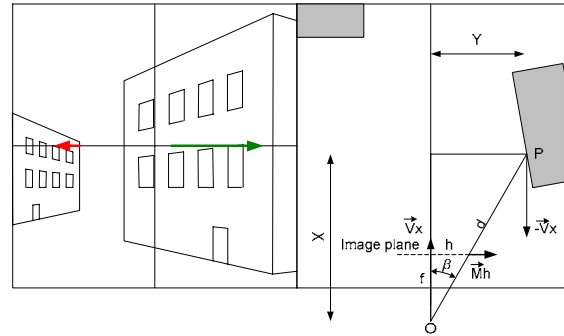


Fig. 7. The UAV in point O perceives obstacle P at the distance d with relative angle β . During a forward translation, any object in the environment is perceived to be moving away from the center of the image (left figure). The obstacle on the right being closer than the one on the left, its horizontal motion is greater and the UAV must turn left.

Furthermore, the high-level controller is endowed with a second reflex that allows the robot to avoid hitting a wall directly in front of it, a situation in which lateral optic flows are equal on both sides of the robot. This reflex calls upon an estimate of the time-to-contact τ , a rough approximation of which is sufficient to prevent the UAV from crashing [8].

If P designates a perceived point, the vector r , projection onto the image plane and the vector M_r the motion of vector r , the Eq. (3) gives the relation between τ and optic flow:

$$M_r/r = 1/\tau \quad (3)$$

One STN controls the averaging of τ in the frame, and informs about the risks of a frontal collision. If the **MEDIAN** result is lower than a minimum value, then such a collision is likely to occur, and the high-level controller triggers a U-turn. As the UAV reminds blind and does not make any decision meanwhile, the controller used in this function implements a *subsumption* architecture [3] according to which the U-turn reflex has a higher priority level than the balance strategy. The more cluttered the environment is, the slower the robot must fly, and vice versa. The same register **MEDIAN** result is a perception of this fact; the controller adapts a velocity depending on this value, but with a priority to stay on the *flight envelope*. (the state variables domain in which the aircraft must remain in order to be controllable.)

The strategy just described works if the UAV displacement is a forward translation. Unfortunately, the rotational component of the optic flow does not depend on distance and,

therefore, corrupts the measurement of the translational component.

Two possibilities:

- Actually, if the housefly seems to be shortening its turning periods as much as possible, it nevertheless does not give up any control ability and uses the inertial information provided by its halteres to compensate for the rotational component of the optic flow. To implement such a solution, we use the balance strategy according to which the UAV was free to move forward and skirt around obstacles and fusion with inertial information. The motion field is a simple vectored summation:

$$M = M(\text{trans}) + M(\text{rot}) \quad (4)$$

- The human being has another reflex; the nystagmus. When you turn the head, the gaze stays on place till the limit, an other gaze is done in front of head. So, before computing the optic flow, we do a dynamic stabilization with 4 STNs and the problem comes back to the forward translation.

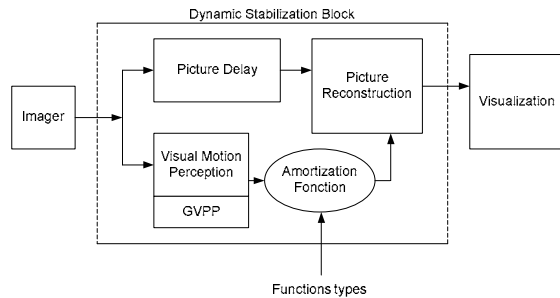
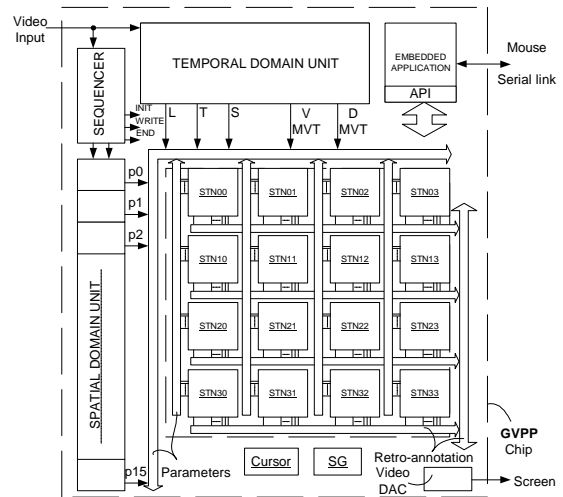
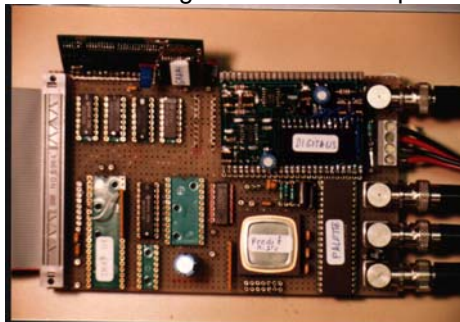


Fig. 8 Dynamic gaze stabilization for FOE

The GVPP chip

GVPP, (generic visual perception processor) is a system on chip, which uses a sequential numeric video as input, and provides application using a specific program inside a Flash memory. A GUI, screen and mouse, is used to emulate application.

Fig. 9. First 1986 chip



7. Perspectives

Since the first chip in 1986 (fig. 9) [10] we have done 7 chips and defined the GVPP methodology. The perception is now extended to other senses as sound with an electronic cochlea as input. Today we finalize a SiP (System in Package). With more STNs, the diversity of applications blows up. The sensors fusion allows complex tasks.

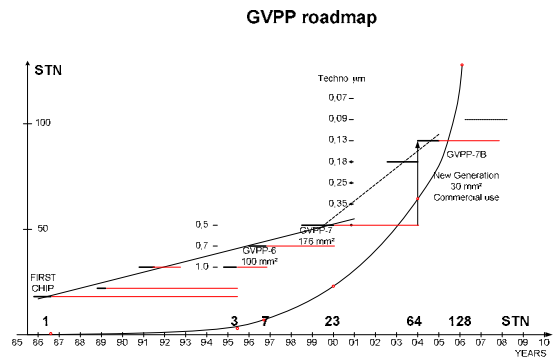


Fig. 10. Evolution nb.STN vs. chip

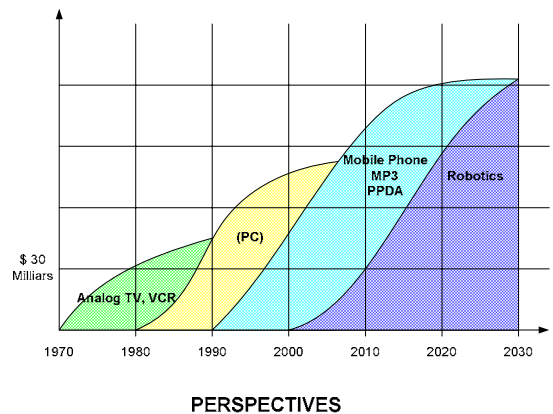


Fig. 12 GVPP in the robotic world

References

1. G. L. Barrows, J. S. Chahl, and M. V. Srinivasan. Biomimetic visual sensing and flight control. *The Aeronautical Journal, London: The royal Aeronautical Society*, 107(1069): 159-168, 2003.
2. P. Buser and M. Imbert, Psychologie sensorielle, 1986, ISBN 2 7056 5944 7.
3. R. A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2:14-23,2003.
4. A. P. Duchon. Maze navigation using optical flow. In p. Maes, M. Mataric, J-A Meyer, J. Pollack, and S. Wilson, editors, *From Animals to Animats 4, Proceedings of the International Conference on Simulation of Adaptive Behaviour*, pages 224-232, Cambridge, MA, 1996. MIT Press/Bradford Books.
5. A. P. Duchon and W.H. Warren. Robot navigation from a gibsonian viewpoint. In *Proceeding of the IEEE International Conference on Systems, Man, and Cybernetics*, pages 2272-2277, Piscataway, NJ, 1994. IEEE.
6. A. P. Duchon, W.H. Warren, and L. Pack Kaelbling. Ecological robotics: Controlling behaviour with optical flow, In *Proceeding of the 17th Annual Conference of The Cognitive Science Society*, pages 164-169, Mahwah, NJ, 1995. Lawrence Erlbaum Associates.
7. A. P. Duchon, W.H. Warren, and L. Pack Kaelbling. Ecological robotics: *Adaptive Behaviour, Special Issue on Biologically Inspired Models of Spatial Navigation*, 6(3/4):473-507, 1998.
8. D. N. Lee. A theory of visual control based on information about time-to-collision. *Perception*, 5:437-459, 1976.
9. L. Muratet, S. Doncieux, J.-A. Meyer, P. Pirim, and T. Druot. Système d'évitement d'obstacles biomimétique basé sur le flux optique. application à un drone à voilure fixe en environnement urbain simulé. In *Proceedings of Journées MicroDrones. CD-ROM ENSICA/SupAero*, Toulouse, 2003.
10. Patent delivery: FR2611063 filed: 1988 08 19, Method and device for real-time processing of a sequenced data flow, and application to the processing of digital video representing a video image.
11. Patent delivery: FR2751772 filed: 2002 11 26, Image processing apparatus and method.
12. Patent delivery: FR2805629 filed: 2002 11 14, Method and device for automatic visual perception.
13. Patent delivery: FR2821459 filed: 2002 08 30, Method and device for perception of an object by its shape, its size and/or its orientation.
14. Patent delivery: FR2843471 filed: 2004 11 05, Visual perception method for object characterization and recognition through the analysis of mono- and multi-dimensional parameters in multi-class computing units and histogram processing
15. N. Franceschini. From fly vision to robot vision: Re-construction as a mode of discovery. In F. Barth, J. A. C. Humphrey, and Secomb T., Editors, *Sensors and Sensing in Biology and Engineering*, pages 223-236. Springer, Berlin, 2002.
16. N. Franceschini. J. M. Pichon, and C. Blanes. From insect vision to robot vision. *Philosophical Transactions of the Royal Society of London B*, 4(4):283-294, 1992.
17. M. V. Srinivasan, Lehrer S. W., W. M. Kirchner, M., and S. W. Zhang, Range perception through apparent image speed in freely flying honeybees. *Visual Neuroscience*, 6:519-535, 1991
18. M. V. Srinivasan, S. W. Zhang, M. Lehrer, and T. S. Collett. Honeybee navigation en route to the goal: Visual flight control and odometry. *The Journal of Experimental Biology*, 199:237-244, 1996.
19. Warren, W. H., Kay, B. A., Duchon, A. P., Zosh, W., & Sahuc, S. (2001). Optic flow is used to control human walking. *Nature Neuroscience*, 4, 213-216

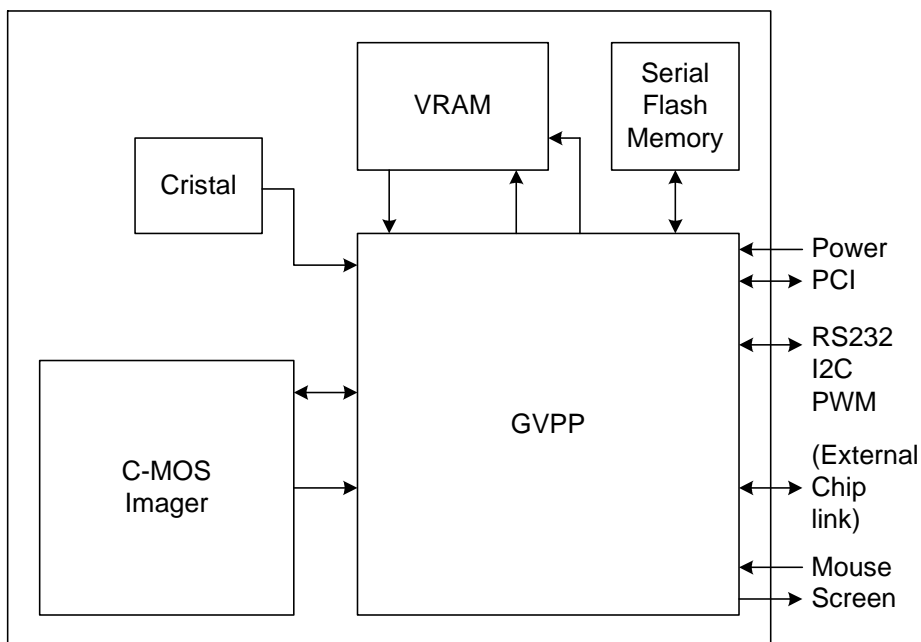
GVPP-7B

Generic Visual Perception Processor



PRELIMINARY INFORMATION

- Array Format (max): 800Hx600V
- Frame rate: 0-100 VGA frames per second progressive-scan
- Interface Mode: Master/Slave
- Data Rate (max): 40 Megapixel per second
- Dynamic Range: 10-bits, 3 chanel
- Parameters: Luminance,Hue, Saturation, Motion (orientation, velocity), Oriented edges, lines, curves, corners
- Computation: 64 STN blocks
- Multi-scales possibilities (optional)
- Multi-chips connections capabilities
- Semi-Graphic inteface visualisation
- GUI with mouse
- PAL, NTSC, VGA, visualisation
- Internal OS
- C language programmation
- PCI, PS2, I2C, RS232, PWM
- 0.3 Watt, 3,3Volts for PAL/NTSC format
- -40,+85 C
- Complete application in one SiP 1"x1": C-MOS Imager, GVPP, 10 Mbits VRAM, 1 Mbits Serial Flash Memory,27 MHz Cristal.



Rev C: 20/03/2004 P.P.