FUTURE VISUAL MICROSENSORS FOR MINI/MICRO-UAV APPLICATIONS

G. L. BARROWS

Centeye, Inc., 6 Logan Circle NW Suite 5, Washington, DC 20005 USA.
E-mail: geof@centeye.com

New classes of small and micro-sized UAVs, with wingspans on the order of meters and tens of centimeters, respectively, present interesting challenges to the field of autonomous flight enabling sensing and control technologies. There is currently a desire to develop a sensor/control suite that will allow such UAVs to fly through complex environments, such as in an "urban canyon" or underneath a forest canopy, at altitudes of just meters above the ground. The development of such capabilities requires new approaches for perceiving the environment. There is increasing interest in borrowing ideas from flying animals such as insects, which are able to fly through such environments with high reliability. This has led to the development of optic flow sensing techniques that currently are able to provide such capabilities as altitude control and terrain following. However, more difficult tasks such as flying in the urban canyon or in a forest require advances in image processing that allow obstacles to be reliably detected by a machine vision package weighing tens of grams, *including* all optics, hardware, and software. A blueprint for such a visual sensor is proposed that makes use of anticipated developments in microelectronic technology. With disciplined "best engineering practices", Cellular Nonlinear Network (CNN) techniques can make significant contributions to the development of such sensors.

1 Uninhabited air vehicles (UAVs), their applications, and challenges

Uninhabited air vehicles (UAVs), or "drone airplanes", are a class of robotic platforms that are becoming increasingly important at the start of the 21st century. Such UAVs were originally designed to perform military missions deemed too risky for a human pilot. Sample applications for UAVs include remote imaging, surveillance, environmental monitoring, and carrying a transponder or repeater for wireless communication. Many of these applications envision UAVs that are able to operate with reliable partial or full autonomy. The desire to provide UAVs with such autonomy has led to the development of reliable autopilots incorporating GPS (global positioning service) receivers, advanced inertial measurement units (IMUs), radar, and control algorithms to allow these UAVs operate for hours at a time without significant human intervention.

Smaller classes of UAVs are becoming increasingly important in both the military and commercial spheres. Such "mini-UAVs" and "micro-UAVs" have wingspans on the order of one or two meters and ten or twenty centimeters, respectively. There is a strong desire to provide such mini- and micro-UAVs with the same autonomy as their full-sized brethren, which has stimulated the development miniaturized sensors and autopilots. Typical mini- and micro-UAVs have payload capacities on the order of several kilograms and tens of grams, respectively. For applications in which a small UAV can fly at a high altitude, the appropriate technologies have been developed that are essentially incremental improvements over existing technologies. However, some proposed applications for these smaller UAVs require that they be able to fly close to the ground, with altitudes measured in meters or tens of meters. At such altitudes, complex terrain features such as hills, buildings, and trees need to be reliably detected and avoided. Current sensing and control approaches are no longer sufficient for such capabilities. Primarily this is because GPS is not sufficiently accurate for such low-level flight, and radar technologies are not mature enough to provide adequate depth perception in the reduced payload capacity of such UAVs. Therefore two challenges are presented:

1. How do we design a package of sensors and control algorithms that allow a UAV to perform its mission with a high degree of reliability?

2. How do we construct such a package so that it can fit within the size and weight budget of a small UAV?

2 Biomimetic visual sensing and control

The desire to be able to fly through such complex environments has led researchers to take inspiration from how animals, especially insects, are able to do the same. Techniques developed with such inspiration are often referred to as “biomimetic”, because behaviors and computational structure observed in biology are essentially mimicked in man-made hardware or software. In summary, insects make heavy use of vision, especially optic flow, for perceiving the environment. Optic flow is essentially the apparent visual motion experienced by an insect (or anything that “sees”) as it travels through an environment. Objects that are close will tend to appear to move faster than objects that are far away, and objects with which the insect are on a collision course will tend to appear as if they are rapidly increasing in size. Numerous experiments on insect behavior suggest that insects use relatively simple heuristics to map an optic flow pattern to a behavior, thereby allowing the insect to achieve a goal (fly to a desired location, land, etc.) while avoiding undesired collisions. The literature on how animals use optic flow to perceive depth, as well as how robots can be equipped with similar depth perception, is extensive. However there is still the challenge of visually perceiving the environment, using optic flow and other cues, using an implementation that is compact enough to fit onto a UAV.

In spite of these difficulties, there are some early successes that serve as a proof-of-principle that optic flow and related visual cues can be used to allow a UAV to fly close to the ground. In one set of successes, a research group led by Prof. Mandyam Srinivasan and Dr. Javaan Chahl at the Australian National University (ANU) was able to demonstrate terrain following in a 2-meter fixed-wing aircraft and both altitude control and hovering in a 2-meter rotary wing aircraft using optic flow as the main method of depth perception. In another set of successes, the author was able to demonstrate altitude control, terrain following, and obstacle detection in a fixed-wing aircraft using 10-gram optic flow microsensors. These microsensors were fabricated with a CNN-like mixed-signal vision chip and a PIC microcontroller. In both cases, aircraft altitude was measured from the optic flow in the downward direction. These successes are very encouraging but currently do not exhibit the reliability required for practical UAV applications. This is certainly understandable given that even the best CPUs available today do not have the processing power of an insect’s or mammal’s nervous system. For example, it can be said that the computational power of the author’s current optic flow sensors is five or six orders of magnitude less than the computational power of a drosophila’s (fruit fly) visual system. Thus, the following question is posed:

3. How can visual information processing be performed such that a complete sensor performing teraops or petaops of useful image processing is squeezed into a ten-gram package?

3 A blueprint for future visual microsensors

To answer difficult questions such as the three above require a combination of both incremental and disruptive innovation and disciplined practicality. To answer the above questions, the author first makes a few claims that, for the purpose of discussion, will be accepted as axiomatic. From these claims, a blueprint for the development of a "future generation" visual microsensor is presented in the next section.

3.1 Claim 1: Moore's Law will continue for at least the next 5 to 10 years, and provide increasingly capable off-the-shelf hardware

The consensus of much of the semiconductor industry is that Moore's Law, the statement that both the number of transistors within a unit area and the switching speed of these transistors will double every 18 month, will continue for at least the next ten years. In choosing architectures and capabilities for future visual microsensors, it is necessary to develop methods that appropriately scale with Moore's Law. Below is a list of some current digital hardware, their approximate performance specifications, and how these specifications would scale with Moore's Law. The implications are staggering.

DSP chip:
Year 2002: 500M flops throughput, 500kB on-chip RAM
Year 2005: 2G flops throughput, 2MB on-chip RAM
Year 2011: 50G flops throughput, 32MB on-chip RAM

Field programmable gate array (FPGA):
Year 2002: 100k logic blocks implementing 8M gates, 400MHz clock rate = 3.2 peta-gate operations per second
Year 2005: 400k logic blocks implementing 32M gates, 1.6GHz clock rate = 51 peta-gate operations per second
Year 2011: 5M logic blocks implementing 500M gates, 25GHz clock rate = 12.5 exa-gate operations per second

Off-the-shelf CMOS imager:
Year 2002: 1.3 Mpixels at 500 frames/sec
Year 2005: 5 Mpixels at 1000 frames/sec
Year 2011: 80 Mpixels at 4000 frames/sec
(Note: resolution can be traded for frame rate)

When choosing a sensor architecture, it is not only reasonable but essential to "plan ahead" for future computational throughputs.

3.2 Claim 2: Massively parallel digital processing is no longer "exotic" but mainstream, and will become even more important in the future.

Massively parallel computers were once a topic of research, but in recent years they have become mainstream and very powerful. In the 1980's parallel computers were multi-million dollar devices purchased by large institutions. In the late 1990's, the introduction of the MMX instruction set by Intel brought parallel processing to mainstream CPUs. Meanwhile advanced graphics rendering chips were being designed with highly parallel architectures to render increasingly detailed images. The most recent graphics chips from NVidia have a computational throughput of one trillion operations per second. These chips are used in the Microsoft X-Box gaming console, which implies that teraops throughput devices have reached the consumer marketplace. If one accepts the crude assumption that image understanding is in some ways an inverse problem of image rendering, then it is not unreasonable that single chip processors for powerful, real-time machine vision are feasible.

One should especially note that by using parallel processing, the Moore's law growth rate is doubled: The number of transistors doubles every 18 months, as does the clock rate, so the throughput actually quadruples every 18 months (or equivalently doubles every nine months). This is about 13 doublings every decade, yielding a throughput increase of 10,000 every decade! It should be noted that this benefit is achieved only if parallel architectures are used, and that the complexity of each processor node remains constant over time.

Consider a "hypothetical" insect brain with 1 million neurons. Suppose each neuron takes inputs from 10000 other neurons. Suppose that the "clock rate" of the insect is 1 kHz, to reflect that faster neuron responses occur in the millisecond time regime. Finally, let us add a multiplier of 10 to account for the computations performed within the dendrite of a neuron. We thus estimate that the computational throughput of the insect brain is $10^6*10^4*10^3*10 = 10^{14}$ or one hundred trillion computations per second. This is only two orders of magnitude greater than the above-mentioned NVidia chip.

Since much of image processing consists of repeating the same instruction over and over again on different data, it makes sense to use parallel architectures in future machine vision systems. Of course, there remains the challenge of knowing what algorithms to implement in such a processor for a given application. (e.g. A hundred trillion useless operations per second is generally still useless.) This is where "best engineering practice" becomes important.

The above claim will be obvious to the CNN community.

3.3 Claim 3: There is not a shortage of models/algorithms for image processing. However, there is a shortage of effective, well-engineered implementations.

An interesting trend has been happening in neuroscience: When a neuroscientist proposes an explanation of a neural circuit, it is no longer considered sufficient to publish a qualitative explanation of how the circuit works. Instead, mathematical or quantitative models are included in the publication. Thus there is a proliferation of biologically inspired algorithms that can be used to create new biomimetic visual sensors. Likewise, the field of image processing is a mature field with an extensive knowledge base from which ideas can be taken. Together, this all implies that there is an extensive set of models, algorithms, and computational structures that are ready for implementation. The author credits Professor Dan Hammerstrom of the Oregon Graduate Institute for providing this observation.

It is the experience of the author, and of every other engineer who is actively engaged in solving real-world problems, that effectively implementing an algorithm into a hardware system (whether as hardware, firmware, or software on a processor) is a non-trivial task. This task requires a level of skill, creativity, diligence, and wit that is on par with that required by biologists who are studying neural systems. A novel algorithm (such as a biologically inspired one) generally cannot "directly encoded" into hardware and be expected to perform optimally. It must be modified, adjusted, tuned, and effectively reengineered into a variant that matches the available hardware medium (whether mixed-signal VLSI or DSP chips) and addresses the specific problem to be solved. Much less work has been performed figuring out how to implement the above algorithms into hardware to solve real-world problems.

Therefore, the author believes that in any task to develop new visual microsensors for UAV (or other robotic) applications, including those using biology as a source of inspiration, the majority of the effort should be directed at implementing such sensors for the purpose of implementing specific sets of capabilities in specific types of robotic platforms. Without such discipline, the sensor will have limited value even if teraflop throughputs are achieved. This is arguably the most controversial claim presented in this paper.

3.4 Claim 4: Better performance can be achieved through the use of a heterogeneous set of complementary algorithms and structures.

In purely redundant systems, multiple copies of the same structure operate on the same data to obtain a result. This way if one of the structures fails, others can take over. In heterogeneous systems, a number of different structures operate on the same data, employing related but somewhat different and complementary functions. This results in an increased diversity of ways to understand the image data and robustly interpret it.

Heterogeneity is a fundamental characteristic of biological systems. No creature relies upon one sensing modality to perceive the world. Even within a single sensory modality, numerous complementary computational structures operate on the same sensor "data" to provide a variety of ways of interpreting the environment. Such heterogeneity would clearly be useful for optic flow sensing. For example, multiple optic flow algorithms operating on the same data can "cross-check" their results, enabling a more robust answer. Likewise, multiple pattern recognition algorithms can increase the number of "ways" to interpret the visual scene. Finally, the results of pattern recognition algorithms may assist in the interpretation of optic flow measurements, and vice versa.

3.5 Claim 5: Future visual microsensors could benefit from the use of advanced focal plane arrays having on-chip processing and/or multiple interlaced arrays

One clear implication of Moore's law, as shown above, is that the decrease in transistor size allows more pixels to be implemented per unit area. An increase in resolution has some benefits for some imaging applications, but is not necessarily the best for all applications. Higher resolution images contain more information that must be processed. Excessive resolution can carry an overhead penalty that adversely affects the performance of an algorithm.

An alternative is to keep the pixel density size constant but make the pixels themselves more capable. This is an appropriate path to take when an imager has sufficient resolution to allow a task to be performed, yet there is room for improving the algorithm by increasing the "quality" of the information within the pixels. There are many techniques for enhancing pixel performance by performing spatial and temporal operations to bring out certain qualities of an image that are substantially more difficult to extract once the image has been digitized. For example, temporal high-pass filtering tends to preserve visual motion information, while some spatial filtering methods can increase the sensitivity of the imager in low-contrast environments or increase the dynamic range of the imager. Such "focal plane processing" embedded in "vision chips" has been the subject of research in mixed-signal VLSI, including within the CNN community. However the implications of Moore's law are that such imagers will soon have sufficient resolution to be practical for wider sets of problems.

This claim is obvious to members of the CNN community.

3.6 Claim 6: Future visual microsensors could benefit from sensitivity to hyperspectral information

Many animals are able to see color, and do so for a good reason- color provides additional information about the visual field that is not obtained from intensity alone. Depending on the application, it make sense for the focal plane of an imager to be able to sense color, and perhaps even enhance color to detect certain types of objects. One can even move further in this direction by implementing the ability to sense beyond the visual spectrum well into the ultraviolet range and/or down into the infrared range. Such hyperspectral sensitivity would allow the sensor to be useful in a variety of weather conditions as well as day or night, and allow visual sensing capabilities beyond that of animals.

4 Proposed canonical sensor

The figure below depicts a canonical visual microsensor that the author believes could serve as a blueprint for the development of next-generation visual microsensors. With a combination of discipline and creativity, it should be possible to fabricate the complete sensor (optics, hardware, and software) in a package weighing on the order of ten to twenty grams.

4.1 Imager

The imager is essentially a focal plane array with photoreceptors and mixed-signal processing. Ideally the focal plane array has hyperspectral sensitivity. The focal plane array performs pre-processing on the image to enhance spatial or spatio-temporal features. Two image sequences are outputted from the imager: One image sequence has a high resolution and a low frame rate, and is suitable for machine vision tasks such as pattern and shape recognition. The other image sequence has a low resolution but a high frame rate and is suitable for measuring optic flow. A CNN-type processor is clearly a good candidate for such an imager.

4.2 Motion sensing / optic flow

Embedded in one or more FPGAs are a number of algorithms for performing motion sensing e.g. optic flow computation. These algorithms take as input the low resolution / high frame rate image data. Several different optic flow algorithms or structures are implemented that estimate the visual motion in complementary ways. This follows the philosophy of heterogeneity, described above in Section 3.4. Some algorithms are simple elementary motion detectors (EMDs) that sense motion in a tightly constrained part of the visual field, while other algorithms track the motion of visual features, such as edges, across larger segments of the visual field. Some motion sensing algorithms may be biological in inspiration, while others may be taken from conventional image processing.

Other algorithms fuse information from these complementary algorithms to arrive at a robust estimate of the optic flow. This section also takes as input information from an inertial sensor (such as a gyro), so that optic flow due to self-rotation can be separated from optic flow due to obstacles. Implementation in an FPGA allows the use of massively parallel structures that yield gigaop or (soon) teraop or even petaop performance in a small package that is quickly reprogrammable and off-the-shelf. In the author's opinion, a combination of digital CNN structures and other computational structures is appropriate for this section.

4.3 Static image processing

Embedded in yet other FPGAs, or on the same one that performs motion sensing, are algorithms or structures that perform tasks generally called "pattern recognition" or "image understanding". These algorithms process the high-resolution and low frame-rate image data. Low-level algorithms pick out features such as edges, curves, or wavelet patterns, while other algorithms use this information to identify possible objects or targets. In the author's opinion, a combination of digital CNN structures and other computational structures is appropriate for this section.

4.4 High-level processing

Embedded in a DSP-type chip are algorithms that extract higher level features from both the optic flow and static image processing results to arrive at a higher level understanding of the visual field. Information from optic flow can be used to enhance the understanding of pattern recognition, and vice versa. For example, if a region of fast optic flow is detected in the same region of the visual field that a vertical pole-like structure is detected, one can reasonably infer that the pole is much closer to the sensor than the background texture. Such higher-level processing is more complex in nature and less easily implementable in a massively parallel architecture. It is for this reason that a DSP-like chip is used instead of an FPGA. This chip also performs I/O to interface the sensor with the outside world. CNN techniques are not appropriate for this stage.

Alternatively, it may be possible to insert all processing onto a single die using recent FPGA products which have one or more CPU cores embedded in the FPGA gate fabric. The motion sensing and static image processing structures could then be embedded within the FPGA fabric, while the high-level processing can be performed in the CPU core.

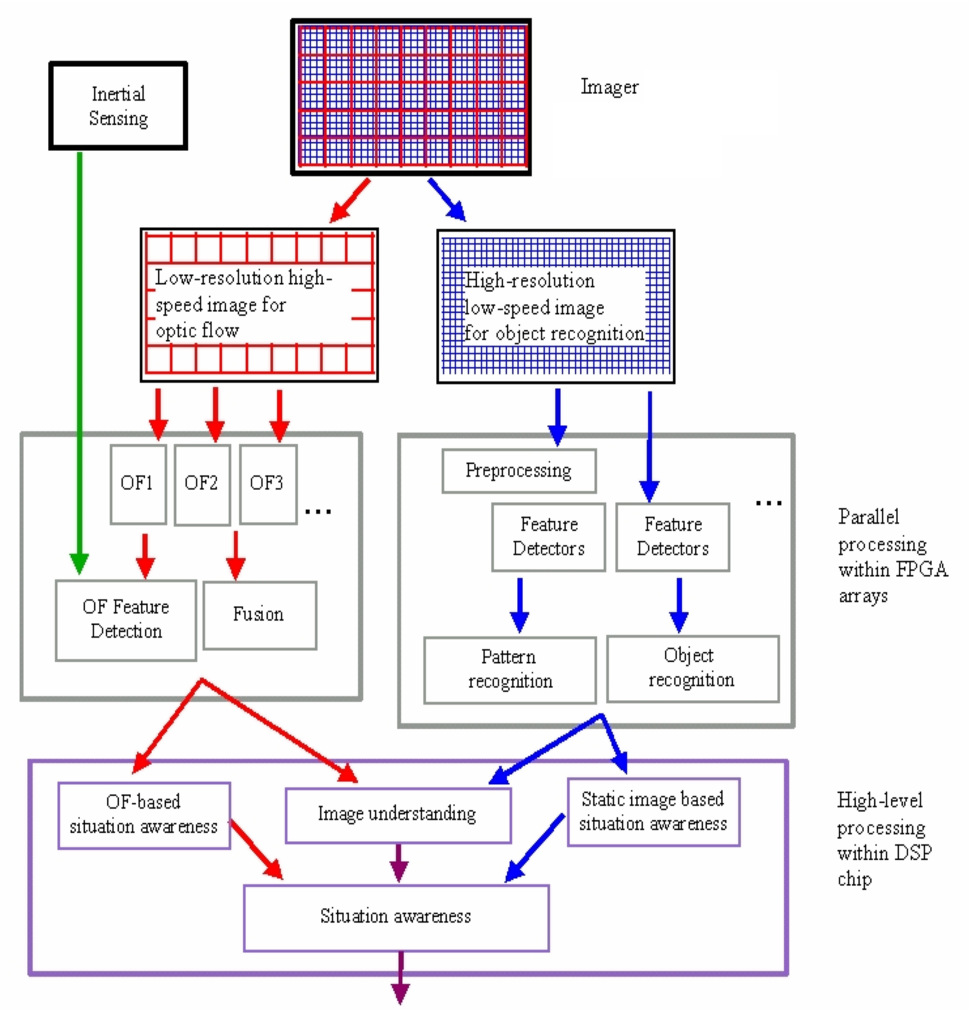

Figure 1: Canonical visual microsensor

5 Possible capabilities to be pursued and implementation issues

With an appropriate blend of creativity, good engineering discipline, and firm commitment, the author believes that the following visual sensing capabilities are possible using the above canonical sensor:

- Fly through an urban canyon, in between and among buildings and other urban structures, without collisions.
- Fly through a forest, underneath the canopy, without collisions.
- Identify and detect visual targets to chase or track (note- the ability to sense color or infrared may be useful for this task).

In order to succeed, the author suggests that the following methods be used in the development of such a sensor:

- From the beginning, develop algorithms and computational structures so that they work with real-world imagery.
- Make integration with a real UAV platform and flight testing a part of the research effort from the beginning. Much can be learned from such flight tests. A typical remote control "model airplane" is an appropriate inexpensive platform.
- Be pragmatic in the selection and implementation of different algorithms. For example, use CNN techniques and other techniques where each is the most appropriate.
- Make a firm commitment to demonstrating a capability, even if initially in a reduced form (e.g. altitude control rather than obstacle avoidance), but always in a real environment. In addition to being extremely educational, such demonstrations serve as a "litmus test" to help evaluate progress, and clearly communicate one's successes to the outside world.

Acknowledgement

The author thanks the Naval Research Laboratory (NRL) and the Defense Advanced Research Project Agency (DARPA) for their support of his research efforts.