Integrated Architectures for Computer Vision Sensing

Denis Poussart, Marc Tremblay and Abdel Djemouai
Laboratoire de Vision et Systèmes Numériques
Département de génie électrique
Université Laval
Québec, Canada

Abstract

The area of sensing for computer vision, especially when it includes range or shape, is rich and diversified, but does not yet provide satisfactory solutions for many applications. The possibility of blending the processing resources of VLSI within the focal plane of optical transducers will lead to sensors with improved characteristics and wider applicability.

Résumé

Le domaine des capteurs de vision numérique, notamment ceux à caractère tri-dimensionnel, exploite des approches conceptuelles variées. Cependant nombre de celles-ci demeurent peu satisfaisantes sur le plan des applications. La possibilité de bénéficier de la capacité de traitement de cirtcuits intégrés à grande échelle, surtout si celle-ci peut être intégrée au sein même du capteur, permettra de réaliser des systèmes de vision plus efficaces et plus polyvalents.

Keywords, mots clé

Sensors, focal plane processing, 3D sensing, VLSI, capteurs de vision, traitement intégré.

Introduction

The general field of sensing - abstracting portions of the physical world into information and signals - implies a wide range of architectural considerations. A case in point is the area of sensing for computer vision. In the quest for "intelligent" systems, i.e. processes which can perceive, reason and act, and which could perform tasks with a minimal amount of supervision, the capability to extract suitable information from the surrounding visual environment is a critical requirement. Architectural considerations arise from a need to blend conceptual aspects together with technological capabilities.

Opportunities for focal plane processing

Vision operates through a process of successive refinements, and originates with direct transduction of light intensity and color. This is followed by the emergence of derived quantities such as edges and regions, and, eventually by more complex percepts such as regions, orientations, motion, depth, etc. Sensing has generally been associated with the lowest level, yielding, for instance, raw intensity data from a dynamic scene. However, as silicon technologies have matured, it has become more and more evident that this simple notion of sensing should be extended. Indeed, future, "smart" sensors will, in many cases, be able achieve higher performance by exploiting the synergy of combining transduction *and* (some) processing on the same solid state substrate or on closely associated subsystems. The so-called *focal plane* processing included in their architecture may seek different benefits: for instance it may complement the physical transduction in order to compensate for some of its deficiencies, and improve the accuracy or the robustness of raw data. Alternatively, or additionally, it may provide for the built-in extraction of higher order primitives. We may anticipate that advanced vision sensors will provide a number of parallel output channels, corresponding to varied types, complexities or spatial extent of vision information.

The reasons for this evolution are manyfold and derives from the dominant characteristics of early processing operations and of the personality of microelectronics:

- Processing which is associated with visual transduction, or which follows it closely involves considerable amounts of raw data. These, however,
- exhibit strong geometric coherence and local support (neighborhood operations),
- are often limited to rather simple arithmetic or logical operations,
- display natural parallelism, and
- exhibit extended spatial regularity.

These requirements lead to architectures which optimize communication and data flow. They can blend effectively with VLSI technology, which provides a natural match to the connectionist character above: wires are costlier to implement than transistors, and optimized use of the silicon resource favors architectures with regular structures, and short, concurrent communication pathways. There exist, therefore, compelling reasons for seeking an integration of transduction and processing within the same silicon substrate [HAL87]. Significant further benefit may arise from such an integrated approach. Processing by digital means, typically favored by the need for reliable communication over long bus lines, becomes less desirable. Since an integrated sensor processes information within microns from its origin, analog processing reappears as a favored candidate strategy. Indeed a major strength of natural vision systems stems from a having arrived to a design where physical (and electrochemical) behavior, in itself, corresponds precisely to the required signal processing operations. Designers of next generation sensors will do well to include this strategy in their tool box [MEA89].

Such integrated sensors represent a subset and special case of neural networks which, as a class, have recently emerged as a computation paradigm which is well suited to VLSI implementation [GOS89]. While a key aspect of general networks, however, is their ability to be topologically reconfigured and thus to support learning, adaptation, and heuristic computation [TRL89], integrated sensors, so far,

have rather emphasized deterministic, algorithmic processing, and attempted to exploit parallelism rather than dynamic reconfigurability [MEA89]. It must be recognized that sensing, in itself, brings strong constraints on the spatial organization of a device. For instance, the small cell size of a 2D photosensing array leaves a limited area to the processing resources. For this reason, until technology allows for much denser circuits, we may have to resort to tightly interconnected networks which are implemented as distinct units.

Significant progress has been made toward the development of integrated vision sensors and, for instance, devices capable of edge [MEA89, DUB89], orientation [ALL88] and motion detection [TAN89] have been reported.

In order to examine some of the related issues and opportunities, we will discuss recent work from our own laboratory dealing with smart vision sensors and dedicated processors which are tightly coupled to raw data acquisition.

High performance range acquisition: sub-pixel interpolation and orientation map computation

The capability to acquire 3D data with adequate accuracy, speed, and robustness remains a serious challenge. This critical component for machine vision remains, as of to-day, one of the bottle-necks for numerous industrial or biomedical applications. Implementations and strategies still tend to be largely application-dependent and span a very wide range of concepts and engineering details [POU88]. At the conceptual level, the type of processing required for extracting raw range or 3D shape data varies considerably and might be qualified as

- simple, for instance in time-of-flight, which involves a straightforward conversion of time or phase to distance,
- intermediate, due for instance to complexity or uncertainty related to dense structured light illumination, such as in Moiré, or
- complex, as in the matching and occlusion issues of stereo, or
- even incomplete, as in attempting to extract shape from shading, where the mathematics are underdetermined.

Although future capabilities may allow for the integration of complex operations, at this time acquisition methods which require modest processing are best candidates for built-in, smart optical sensors. For instance, methods of 3D sensing which are based on variants of structured light illumination requires the accurate measurement of the position of the light pattern reflected from a scene, and belong to that class. Sub-pixel interpolation [BLA86], which can provide positioning accuracy beyond the raw resolution of the sensor, exemplifies this level of processing. Indeed, active triangulation ranging is a promising field of application for focal plane-processing.

The principle of active triangulation is straightforward. As shown in Figure 1A, a collimated beam of light with known orientation α is cast upon a scene. The reflected signal is imaged upon a position detector, whereupon range z is simply computed by triangulation. The ability to rapidly and reliably estimate the position of the detected light signal is critical.

This simple geometry, however, does not lead to an optimal use of the resolution.speed - of - response product which characterizes a given detector technology since both the x location and the z range of the target point are simultaneously encoded onto the detector surface. The architecture of the synchronized scanning method [RIO84] recognizes the close interplay between geometry and detector capabilities: as illustrated in Figure 1B, a second scanning device, M2, is introduced in the return path and its deflection angle is slaved to α. It can then be shown that the detector maps z range directly. This property has the potential of yielding excellent 3D sensing and prototypes with 300 x 300 resolution at close to video rate have been demonstrated. The achievement of high accuracy of range data must, however, take into account the physics of optical reflection from an arbitrary target. In particular, if the local reflectance properties of the scene are not spatially homogeneous within the effective cross-section of the scanning beam, the light distribution on the position detector is modified, and significant range errors can occur. We have modelled this problem in some detail and have proposed correction strategies [SOU90].

Compensation requires that a detailed profile of the light distribution be available. On the one hand, this condition rules out the use of devices such as lateral photodiodes which, by virtue of their differential geometry, have the otherwise advantageous property of fast, direct read-out of the effective central position of an arbitrary light distribution. CCD sensors, on the other hand, can provide the detailed profile of optical input which is needed for correction, but at the cost of a relatively slow, sequential read-out, and external processing circuitry. This is a significant constraint for dense range data since a complete read-out of the CCD sensor must be repeated for each target point.

Figure 2 shows the layout and floor plan of a prototype focal plane processor which is currently in development in our laboratory and which supports sub-pixel position sensing with direct digital read-out.

Optical input from a single mode distribution is imaged onto a linear array of 34 NPN floating-gate phototransistors (size of 105 x 112 μm with inter-pixel gap of 27 μm) The corresponding linearized photocurrents are submitted to a parallel comparison with a common threshold which scans the amplitude range in order to seek the location of the pixel with maximum intensity Imax. This address is made available as the 5 bit, integer component of peak position (means are provided to account for instances where two cells record identical intensities). A gating circuit feeds I_{max} and intensities I_{max+1} and I_{max-1} from the adjacent cells on either side of the maximum (hence the reason for 34 pixels to accommodate 32 discrete positions without edge effects) to a compound A-D converter. This subsystem, in turn, computes the quantity

$$\frac{\left|I_{max+1}+I_{max-1}\right|}{4I_{max}-2I_{max-1}-2I_{max+1}}$$

which is outputted as the fractional binary component of position.

Simulations have shown that for a light distribution with gaussian profile of width σ, this simple algorithm can yield

suitable estimates of the (sub-pixel) position of the peak. Accuracy can reach a few percent when the effective width σ spans several pixels. Of course the spatial uniformity of the photosensitivity of the cell array is a critical element. Since identical transistors may exhibit significant variations, the device includes means for automatic gain compensation: as I_{max}, I_{max-1}, and I_{max+1} are fed to the interpolator, each current is finely adjusted, in 4% increments, through a 4-bit gain normalization channel which is driven from external ROM. This reference is addressed from the integer peak position data, and its magnitude is determined and stored during an initial gain calibration procedure.

Versions of this device are being implemented in the CMOS3 technology through the fabrication facilities of the Canadian Microelectronics Corporation [CMC89]. We anticipate computation time in the range of 5 μsec when the design is refined. 1-dimensional integrated sensors such as this one are suitable for the eventual inclusion of fairly complex, but regular, operations since the second dimension of the silicon area is available for processing circuitry. As denser fabrication techniques, such as the 1.2 micron CMOS4S process of the CMC, become available, we can expect that higher order processing, such as strategies for compensation of optical profiles distorted from reflectance variations, can be integrated as well.

The BIRIS sensor [BLA85a,b], which exploits defocusing in a 2D intensity image to infer range data is another example of an architecture where the optical properties and signal processing capabilities are efficiently blended to yield a 3D sensor with remarkable qualities [BLA89]. The processing requirements are specific, but fall in the same general category as those discussed above: pairs of lines, or more complex optical patterns must be very precisely located at the focal plane. Real-time versions of this sensor using external special purpose image processing boards have been demonstrated. Again, we feel that VLSI technology with the flavor described above will eventually be able to process the intensity information within the camera itself, thus leading to a sensor with exceptional small size and mechanical robustness, well suited to a range of applications in demanding environments.

Because of the topological constraint of 2D arrays, the complexity of processing performed in immediate proximity of the optical transduction is limited (see below). However, it is possible to design subsystems which are to be used with very close, dedicated communication with an acquisition array, so as to blend with considerable synergy and to lead to what can be effectively considered as a smart sensor. For instance, we have described special purpose VLSI circuits designed to rapidly compute the direction of local orientation from a range map [BLN87]. We can envision such processors, operating on a local base (say over a grid of 5 x 5 voxels) and being included within a storage memory with built-in bi-dimensional addressability [TER87] and high speed sequential scanning. We envision future computer vision sensors as devices with multiples data modalities - for instance, intensity, range, and orientation -, and within which a sizable fraction of the manipulations required for early segmentation are efficiently and transparently performed.

2D architectures for feature detection and tracking

Work toward "artificial retinas" has been strongly oriented toward the on-chip integration of specific image processing. Because of the small space available to implement arithmetic or logical operators directly attached to each pixel of a dense 2D arrays, current technologies imply simple operators with great spatial homogeneity. Analog processing is preferable, especially when it is coupled to and designed to exploit the natural 2D response of interconnected mesh.es For instance, resistive connections between pixels effectively synthesize useful convolution kernels and a basic strategy consists of implementing circuits whose direct physical behavior happens to match the desired algoritm, thus leading to a form of special purpose analog computers [MEA89]. An integrated sensor with real-time edge extraction which we have described [DUB89] is an illustrative example of this approach.

More complex operations, however, especially when they must exhibit variations as a function of the local characteristics of an image, can hardly be implemented in this manner. An on-going project explores the trade-off between the simplicity of basic pixel processing and the complexity and penalty of rapid communication with tightly coupled, but external, processing support.

As shown in the block diagram of Figure 3, the "Multi-port Array Receptor" (MAR) system, which is described in more details elsewhere [TRM89,90], links three units in a closed-loop configuration: *i)* a 2D array of photosensitive elements, *ii)* an analog processing module, and *iii)* a controller unit.

The single photo-currents from the 2D array are retrieved through a set of half transmission gates with individual selection lines organized in a fashion of a multi-port memory with individual data bus. As shown in Figure 5, selection buses are routed at relative orientations of 60 degrees and define an overall tesselation with an hexagonal matrix. This topology allows for simultaneous access to the illuminance data of the selected pixel (intersection of the three selected lines) together with the illuminances of all neighbors located on the three axes of symmetry of the sensor array (corners of the concentric hexagon). This topology yields a natural compatibility with circularly symmetric operations since all pixels located at the corner of a given hexagon have identical radial distances. Furthermore, in contrast with a conventional raster scanner, the MAR sensor can be scanned freely along any of the six possible directions of the underlying hexagonal structure. A unique address and data buses organization provides parallel analog readout of the illuminance data $E(r)$ from rings of pixels of increasing radii r (up to 9 pixel units in the current implementations) located along the main 60° axis and centered around any pixel of interest (POI). Operations to be performed over (symmetrical) regions thus have built-in support.

A major goal of this integrated sensor is to perform real-time edge detection. As stated in [MAR82], it is desirable to perform this operation at several different spatial resolutions. Image segmentation may then proceed though a sequence of gradually finer filters which converge toward a precise extraction of small details of the scene. One purpose of the analog module is to compute the required convolutions by weighing the set of $E(r)$ data by coefficients which

corresponds to sampling the Marr filter for different radii. Spatial zero-crossings of the filtered image determine the location of edges at the selected scale.

In this first prototype we have chosen to implement this precise processing by external circuitry, which, at the moment, can best provides the required combination of wide bandwidth and high input impedance. Several analog convolution modules are used in parallel according to the different spatial resolutions and support the zero-crossing detection. The resulting outputs as well as raw image data may be sampled and converted to a digital form suitable for a host computer, for instance via a DMA channel. These outputs are also fed back to the controller module to allow conditional displacement of the POI.

The relative and random access nature of displacements of the POI within the MAR sensor requires several direction signals. These are generated by the controller module whose block diagram is illustrated in Figure 6. The controller implements an extensive instruction register which defines the operating modes of the camera. The instruction register drives the inputs of a programmable logic array (PLA) which implements a state machine, where specific microcoded instructions are executed.

The main characteristics of the instruction set are outlined in Figure 7. This initial set is sufficient for performing a range of early, low level image acquisition and analysis such as edge finding and tracking or maximum - minimum detection It also provides support for more complex processing such as shape-from-shading computations, local curvatures, local correlation for stereoscopic vision, and sub-pixel inperpolation in 2D or 3D data. With additional interaction with its host computer, the approach described here is suitable for the efficient extraction of higher-level features such as area, shape, moments of inertia, center of mass, motion, etc.

Comprehensive simulations of the convolution kernel and zero-crossing algorithms have been performed and results confirm that this architecture has promising potential. An early implementation of the MAR sensor has been performed using the CMOS3, 3 μ technology. A 4000 x 4000 μ full custom 24 pins chip is currently in queue for fabrication and its l floor plan are shown in Figure 4. This version consists of a 64 x 64 array and yields data for POI with a maximum radius of 9 pixels and a convolution kernel which spans 49 pixels. The design is being ported to the CMOS4S, 1.2 μ technology, which will allow for pixel arrays in excess of 200 x 200 pixels.

Conclusion

This paper has presented an overview of a design approach which aim at exploiting the custom capabilities of VLSI toward the realization of efficient vision sensor. The inclusion of built-in processing capabilities can be used either for enhancing the fundamental physical characteristics of the transducing operation or to extract higher order information to be used by the subsequent recognition and interpretation operations. Many applications of computer vision, especially when range information is required, is still hampered by the availability of sensors with proper characteristics, including (some) "smartness". Sensors architecture which integrate astute acquisition principles with focal plane processing will, in the future, contribute to expand the areas of applications.

Acknowledgments

This work was supported in part by grant A5274 of the Natural Sciences and Engineering Council of Canada (NSERC) and grant 89EQ2830 of FCAR (Province of Québec). The Canadian Microelectronics Corporation (CMC) provided software, hardware and fabrication support. M. Tremblay and A. Djemouai were supported financially by a FCAR postgraduate scholarship and the Algerian government, respectively. We also thank Yanick Tremblay for the computer simulation of the microcode as well as the entire controller unit.

Bibliography

[ALL88] Allen *et al.*, An Orientation-Selective VLSI Retina, *Proc. 1988 SPIE Conf. Visual Comm. and Image Processing*, SPIE, Bellingham, Wash., 1988, 1040-1046.

[BLA85a] Blais, F. 1985. Capteur optique de vision de formes tridimensionnelles pour applications industrielles (in French), MS thesis, Dept of Electrical Eng., Université Laval.

[BLA85b] Blais, F., Rioux, M., Poussart, D., A very Compact 3-D Camera for Robotic Applications, *Proceedings of Meeting on Machine Vision*, Optical Society of America, paper WB4-1, 1985.

[BLA86] Blais, F., Rioux, M., Real-time Numerical Peak Detector, Signal Processing, **11**, 1986, 145-155

[BLA89] Blais, F., Rioux, M., and Domey, J., 1989, Compact 3D Camera for Robot and Vehicle Guidance, Opt. Lasers Eng., **10**, 227-239.

[BLN87] Blanchet, M., Poussart, D., In-Memory VLSI Processor for Computing Local Surface Orientation of 3D Range Images, *Proceedings of Automated Inspection and High Speed Vision Architectures Conference*, SPIE Symposium on Advances in Intelligent Robotics Systems, Cambridge, 1987.

[CMC89] *Guide to the Integrated Circuit Implementation Services of the Canadian Microelectronics Corporation*, Revision 4.0, March 1989.

[DUB89] Dubois, D., Poussart, D., Real-Time Image Edge Extraction with Integrated CMOS Sensor, *Proc. Vision Interface '89*, London, 70 - 76, sept 1989.

[GOS89] K. Goser, U. Hilleringmann. U. Ruekert, K. Schumacher, VLSI Technologies for Artificial Neural Networks, IEEE Micro, Dec. 1989, 28-44.

[HAL87] D.G. Hall, Survey of Silicon-Based Integrated Optics, *Computer*, **20**, Dec 1987, 25 - 39.

[MEA89] C. Mead, *Analog VLSI and Neural Systems*, Addison-Wesley Publishing Co., Reading, Mass., 1989.

[MAR82] D. Marr, *Vision*, W.H.Freeman and Company, 1982.*Computer*, **21**, Mar. 1988, 52 - 63.

[POU88] Poussart, D., Laurendeau, D., 3D Sensing for Industrial Computer Vision, in *Advances in Machine*

[POU88] Poussart, D., Laurendeau, D., 3D Sensing for Industrial Computer Vision, in *Advances in Machine Vision: Applications and Architectures*, J. L. C. Sanz, Ed., Springer-Verlag, 122-159, 1988.

[SOU90] Soucy, M., Laurendeau, D., Poussart, D., Auclair, F., Behaviour of the Center of Gravity of a Reflected Gaussian Laser Spot Near a Surface Reflectance Discontinuity, in press in *Journal of Industrial Metrology*, 1990.

[TAN89] J. Tanner, C. Mead, *Optical Motion Detector, in* C. Mead, *Analog VLSI and Neural Systems*, Addison-Wesley Publishing Co., Reading, Mass., Ch. 14, 1989.

[TER87] Tervo, R., Poussart, D., Sequential/Parallel Instruction Set Extension for Image Processing, *Proceedings IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, 48-50, 1987.

[TRL89] P. Treleaven, M. Pacheco, M. Vellasco, VLSI Architectures for Neural Networks, IEEE Micro, Dec. 1989, 8-27.

[TRM89] Tremblay, M., Poussart, D., MAR: an Early Vision System with Integrated Optics and Processing, *Proc. Canadian Conference on Very Large Scale Integration*, Vancouver, 33 - 40, oct 1989

[TRM90] Tremblay, M., Poussart, D., MAR: an Integrated System for Focal Plane Edge-Tracking with Parallel Analog Filtering and Buit-in Primitives for Image Acquisition and Analysis, to be presented at the *IEEE International Conference on Pattern Recognition*, Atlantic City, june 1990.

[RIO84] Rioux, M. 1984. Laser Ranger Finder Based on Synchronized Scanners, Applied Optics, **23**, 3837 - 3841.

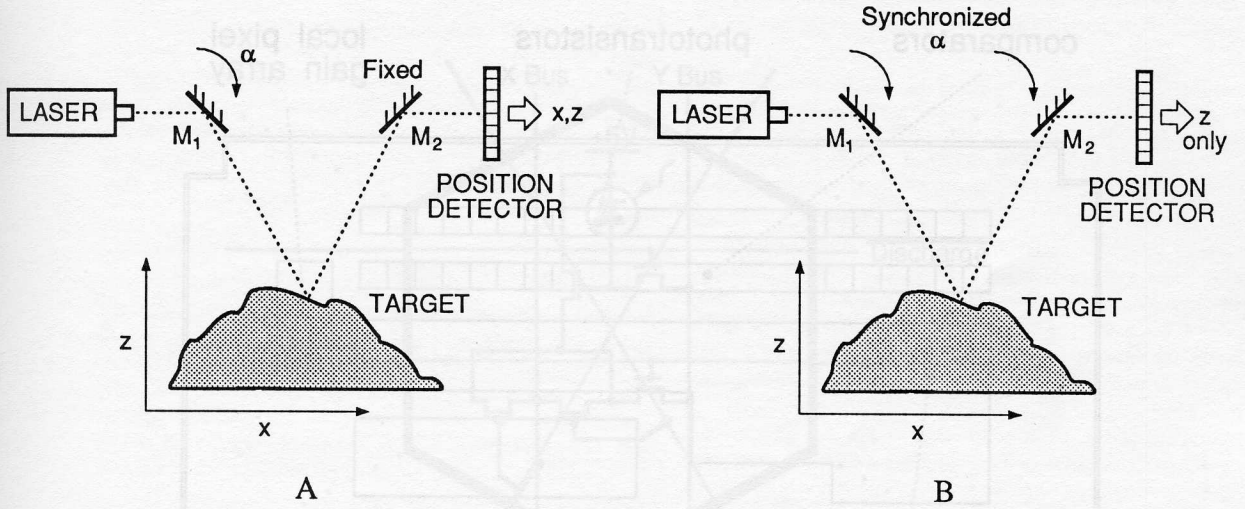
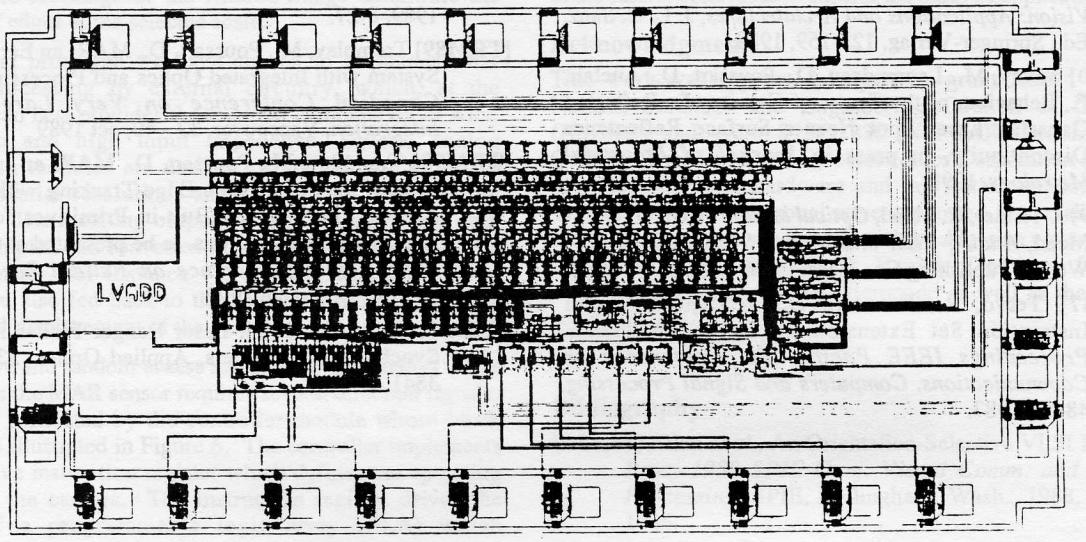


Figure 1. The active triangulation method of 3D ranging (A) is a good example of a generic type of sensor where performance is ultimately limited by the precision and speed of the position detector. In its synchronized scanning implementation (B), the optical arrangements uses two slaved mirrors, resulting in a condition where only range *z* is mapped onto the detector. This important advantage may be further amplified by "smart" position sensors operating at the focal plane



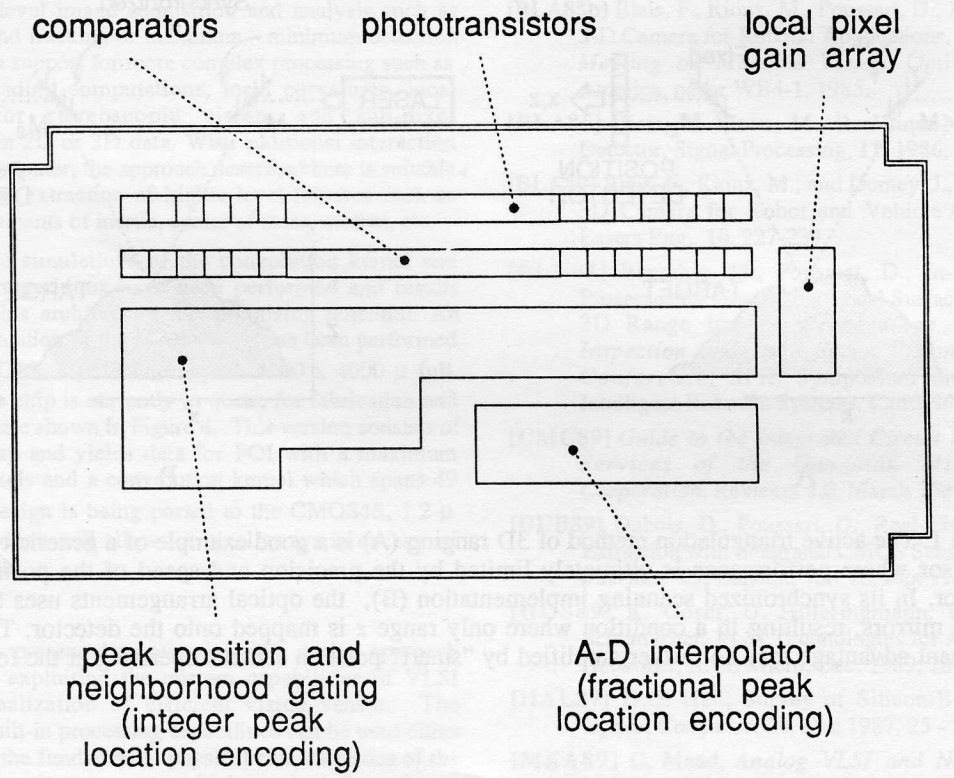


Figure 2. Layout and floor plan of a prototype sub-pixel resolution integrated sensor.

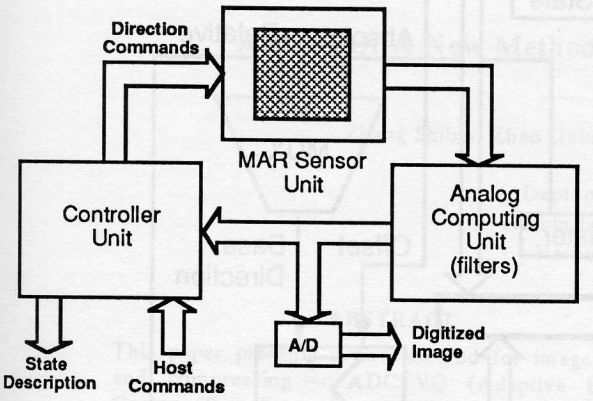


Figure 3. Block diagram of the MAR integrated system

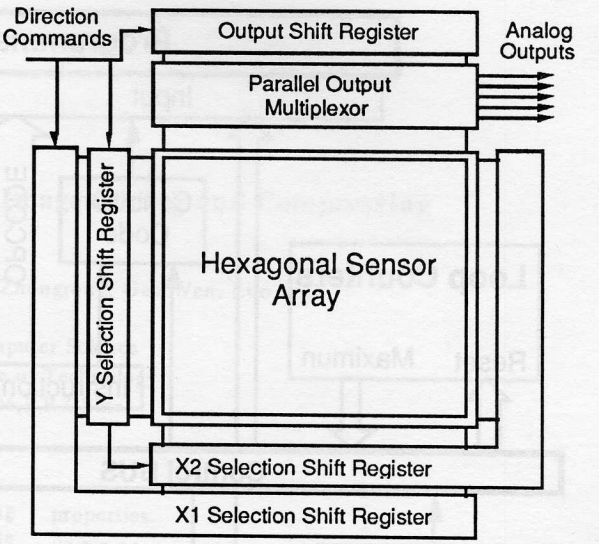


Figure 4. Floor plan of the MAR sensor

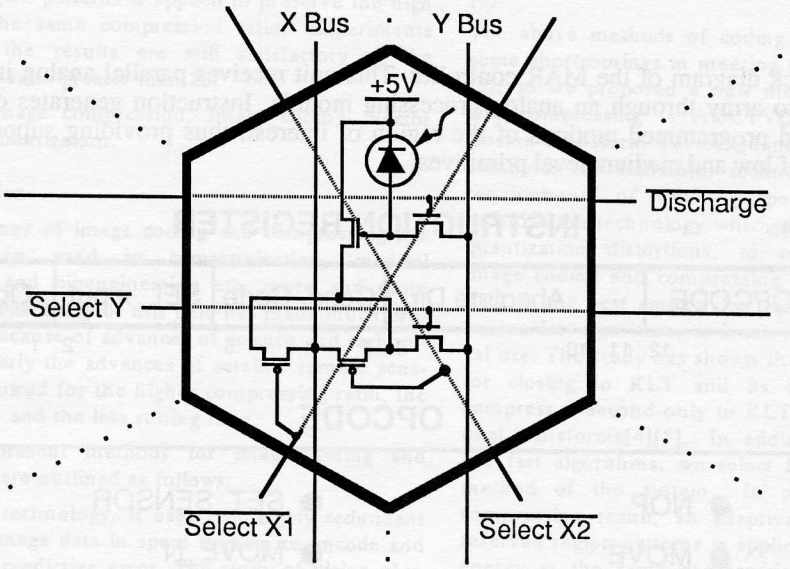


Figure 5. The MAR sensor is built around an hexagonal tesselation. Selection lines determine an activated region consisting of a central pixel surrounded by regions projecting along the main diagonals. The data paths provide parallel illuminance data on several hexagonal paths surrounding the central pixel, and provide the basis for convolution kernels operating at multiple resolutions. Arrays of more than 200 x 200 pixels are being prototyped in 1.2 micron CMOS.

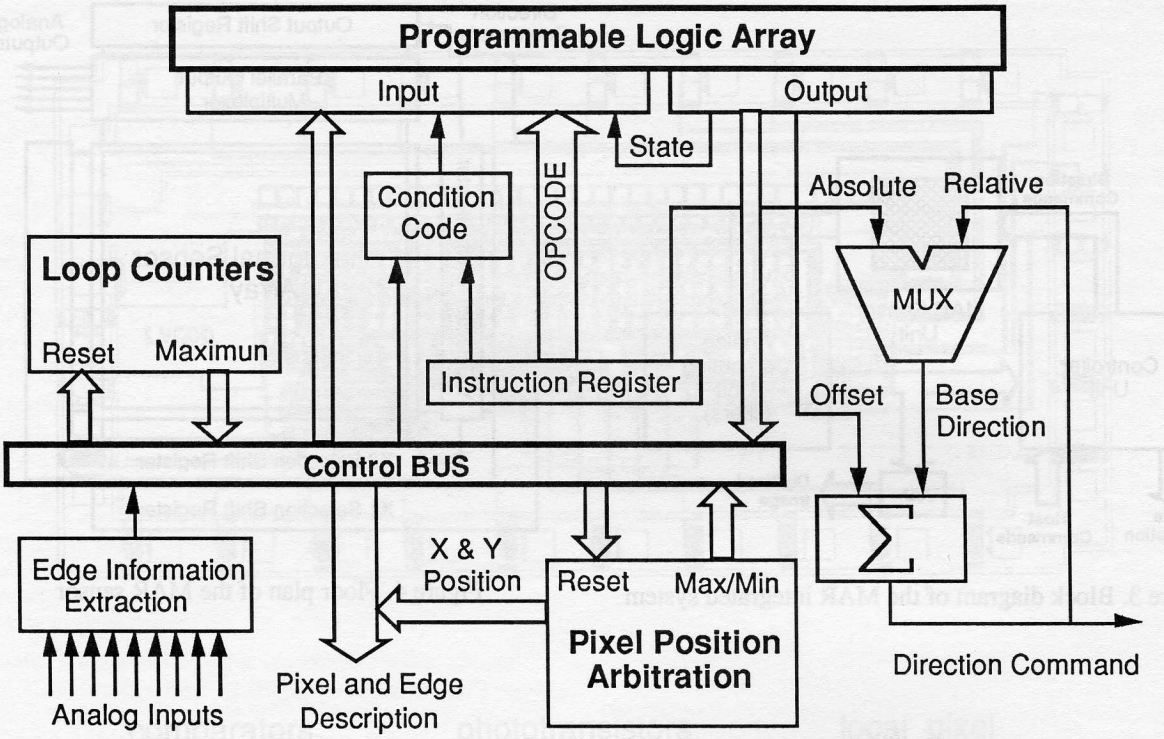


Figure 6. Block diagram of the MAR controller. This unit receives parallel analog information from the photo array through an analog processing module. Instruction generates conditional operations and programmed motions of the region of interest, thus providing support for the computation of low and medium level primitives.

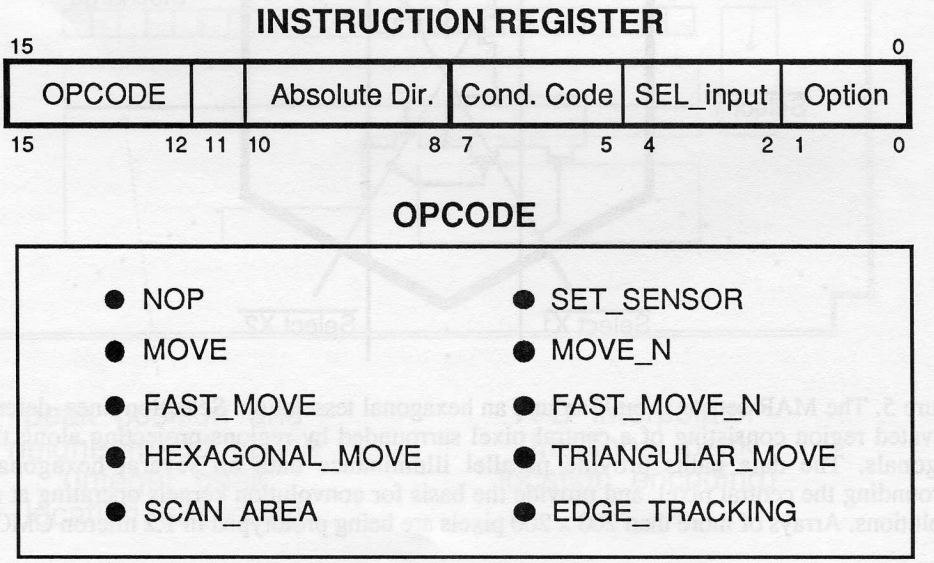


Figure 7. Overview of instruction format implemented by microcode within the controller.