

# A $40 \times 40$ CCD/CMOS Absolute-Value-of-Difference Processor for Use in a Stereo Vision System

J. Mikko Hakkarainen, *Member, IEEE*, and Hae-Seung Lee, *Senior Member, IEEE*

**Abstract**—This paper presents an analog VLSI processor chip with application in a high-speed binocular stereo vision system used for the recovery of scene depth. We have attempted to exploit the principal advantages of analog VLSI—small area, high speed, and low power—while minimizing the effects of its traditional disadvantages—limited accuracy, inflexibility, and lack of storage capacity. A CCD/CMOS stereo system implementation is proposed, capable of processing several thousand image frame pairs per second for  $40 \times 40$ -pixel binocular images. A  $40 \times 40$ -pixel absolute-value-of-difference (AVD) array, a core processor of the stereo system, was fabricated in a  $2\text{-}\mu\text{m}$  CCD/CMOS process. Individual unit cells in the array were characterized and tested. The array functionality was next tested by imbedding it in a computerized stereo system and using both real-scene and computer-generated input image pairs. The system output was compared with full computer simulations for the same image pairs, showing good correlation.

## I. INTRODUCTION

RECENT developments in VLSI machine vision suggest that special purpose analog and digital circuits can be used to achieve real-time processing rates (e.g., [1]–[4]). This work addresses speed and cost of computation as they pertain to the design and fabrication of a high-speed stereo vision system. Previous implementations of stereo algorithms have utilized mainly the massive computational powers of parallel supercomputers. Although these machines perform well as algorithm test beds, their success has been limited in terms of system speed and cost. For example, the 65,536 processor Connection Machine designed specifically for two-dimensional image processing requires several seconds to compute a stereo disparity map for  $256 \times 256$  images using the Marr–Poggio–Drumheller algorithm [5]. A real-time processing rate is not achieved in this case because the underlying hardware does not directly support the communication structure of the algorithm even though computations are performed in parallel (“computation” is used to refer to the actual operation(s) performed by individual processor(s) in the vision system while “communication” either denotes interactions between different/memory cells or refers to external I/O’s).

Manuscript received December 8, 1992; revised March 2, 1993. This work was supported by the National Science Foundation and DARPA under Contract MIP-8814612.

J. M. Hakkarainen was with the Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139. He is now with the General Electric Research and Development Center, Schenectady, NY 12345.

H.-S. Lee is with the Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139.

IEEE Log Number 9209021.

Special-purpose vision circuits and systems are designed for a particular algorithm or a set of closely related algorithms and thus are optimized for maximum computation and communication efficiency. Several digital chips are already available that can perform specific image computation such as linear spatial filtering, pixel-wise addition, subtraction, and multiplication [6] with significant gains in speed and overall cost. However, a full vision system made from a collection of these modules still requires several chips and considerable board area. Novel approaches in both digital and analog array processors try to reduce system size even further. The “smart sensor” approach tackles computation and communication bottlenecks in machine vision algorithms by coupling image data processing tightly with the sensors [1]. At each computational stage, much effort is spent on reducing the data transmission bandwidth. Furthermore, many smart sensor systems use analog processing and push the analog/digital interface (A/D converters) away from the imagers. The stereo vision system in this paper is an example of such a system because its analog processors provide a digital output without using any A/D converters.

This paper is organized as follows. Section II introduces the binocular camera geometry and the stereo algorithm used in this work. The overall stereo system hardware architecture is presented next in Section III, followed in Section IV by a detailed description of the absolute-value-of-difference (AVD) chip, a core system processor. Finally, test results for the AVD chip are presented in Section V.

## II. RECOVERY OF SCENE DEPTH FROM BINOCULAR STEREO

Depth perception is the process of recovering 3-D information from 2-D images. In order to understand how 2-D images can be reliably mapped back to the 3-D world, most researchers have divided the depth perception problem according to the various available cues. This has led to studies in binocular stereopsis, shading, motion vision, and stochastic stereo matching (e.g., [7]–[11]). This work deals with binocular stereopsis and the stereo correspondence problem in particular. Binocular stereo is based on the fact that a scene point is imaged to different relative locations in the two images, as indicated by the projections of point  $P$  in Fig. 1. Stereo vision (depth recovery) is achieved if the difference in relative coordinates, also called the disparity, of corresponding points is determined. Fig. 1 reveals that the corresponding points are constrained by geometry to lie on particular corresponding epipolar lines. The image planes in Fig. 1 are in a coplanar orientation, which causes all epipolar

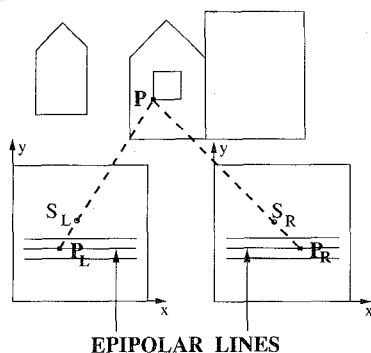


Fig. 1. A coplanar stereo vision geometry showing a scene point  $P$  imaged to left image point  $P_L$  and right image point  $P_R$  ( $S_L$  and  $S_R$  are the camera lens centers).

lines in the two images to be mutually parallel and to coincide with imager pixel rows. This geometry simplifies the search process for corresponding point pairs, since each left image pixel only needs to be compared with a range of pixels on the corresponding row in the right image. Note that the search process would be more complex if the images were not coplanar because the orientation of the epipolar lines would change quite dramatically. In fact, each epipolar line would be oriented at its own unique angle across the image plane. The search for corresponding point pairs would thus require crossing pixel rows as well as possibly keeping track of the orientations of various epipolar lines. This case was considered too complicated for the present implementation. The trade-off for assuming coplanar images is that the fields of view of two cameras do not overlap sufficiently for scenes too close to the cameras. This occurs when the distance to the scene is short enough to approach the interocular distance (= baseline, the distance between the two camera lens centers). In practice, objects a few interocular distances away can be allowed without significant sacrifice in performance (further analysis of binocular camera geometry can be found in [8]).

Despite a simplified camera geometry, the recovery of corresponding points is still difficult because a single pixel in one image can typically match many pixels within a given range of the other image row, giving rise to several false matches. The false matches can be random occurrences, although scene foreshortening (caused by surfaces that slope away from the cameras) and occlusion (due to surfaces seen by one camera but occluded from the other) are notorious for introducing matching problems. Most reported stereo correspondence algorithms locate the best matches by determining "scores" for each match pair through a local support computation process. Best scores are then used to select the most likely match pairs, allowing one match for each left and right image pixel. This process is translated into the following four distinct computational steps.

*1. Raw image preprocessing:* The purpose of this step is to emphasize those features in the input image pair that are considered to provide reliable matches in the next step (this step is often also referred to as extraction of "matching primitives"). Proper preprocessing has been shown to reduce the possibility of selecting a false match over a correct match [5]. For example, emphasizing brightness edges with a

spatial high-pass filter yields much better results than using unprocessed images. An intuitive explanation for this case is that spatial high-pass filtering removes constant brightness levels (spatially "dc"), which can differ greatly between the two images, partly because the cameras look at the scene from slightly different directions. Matching of the high-pass filtered images thus depends on finding equal brightness gradient locations and not on the brightness levels themselves. This eliminates the possibility of bias towards particular brightness levels. It should be noted that many different schemes exist for selection of "matching primitives" and one is usually limited by how much system (hardware or software) complexity can be afforded at this step. In the interest of high-speed analog computation this step cannot be too complex, and thus simple linear filtering seems to be a good option.

*2. Computation of a range of match data for each pixel:* Each pixel in one preprocessed image is compared with a predetermined range of pixels (the maximum expected disparity) in the other image, producing candidate matches. Thus two  $N \times N$  images with a comparison range of  $D$  pixels yield a total of  $DN \times N$  candidate match arrays. Since  $D$  is typically on the order of  $0.1N-0.2N$ , this step introduces a potential speed bottleneck to the algorithm. The "comparison" metric used in this design is the absolute value of difference (AVD), although the square of difference and many other functions could also be used.

*3. Computation of local support scores for all match data:* The candidate match arrays from the previous step are processed further to improve the reliability of best-match selection. Usually neighboring candidate matches supply their individual "votes" or local "support" values that are combined to form a local support score for each pixel. For example, if candidate matches use the AVD metric, local votes are proportional to the AVD at the voting pixel, and support scores are formed by (possibly weighted) addition of all individual votes that come from a predetermined local neighborhood around each pixel. The voting process may occur between matches computed at the same disparity or even between matches at different disparities. Authors differ widely at this step on how local support computation should be done and how a local neighborhood should be defined. This is discussed further below.

*4. Selection of best matches for each pixel:* Using the local support scores a unique best match is selected for each pixel. For example, best matches for AVD-based local support have the smallest support scores. The disparity value that corresponds to the best match is provided as output (this decision process does not provide subpixel resolution). Thus, the result of stereo computation is a map of disparities over the  $N \times N$  images. Since disparity is inversely proportional to distance, scene depth can be calculated from this map.

Many current stereo correspondence algorithms follow these four basic steps. Most schemes are actually quite similar in steps 1, 2, and 4, while the crucial differences can be found at step 3 (local support computation). The Marr-Poggio-Drumheller (MPD) algorithm was compared against the Pollard-Mayhew-Frisby (PMF) [12] and the Prazdny [13] algorithms as reported in more detail in [14] and

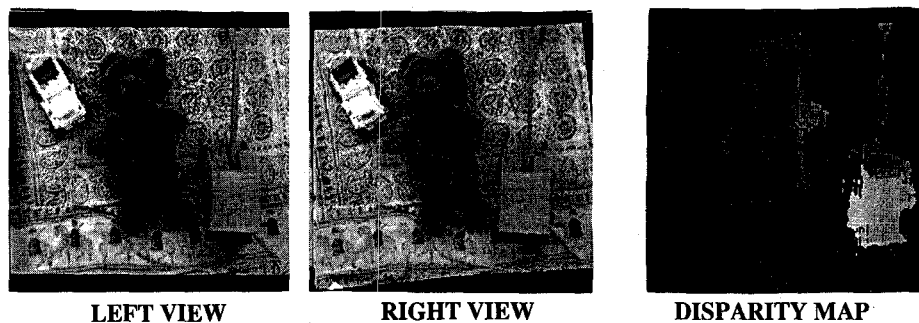


Fig. 2. Input image pair with the resulting MPD stereo system disparity map output.

[15]. All three algorithms are well established in the field, with PMF and Prazdny perhaps achieving better overall accuracy than MPD.

MPD uses spatial low-pass filtering over each of the candidate match data arrays for local support. Each match in this scheme receives votes from other matches at the same disparity only. The local support score is thus formed by the weighted average of all candidate matches (AVD values) in a local neighborhood defined by the extent of the low-pass filter. This has potential for fast computation and requires a small amount of memory space since only one candidate match data array is needed at a time to produce the next local support score array. Thus, local support computation may be conveniently synchronized with candidate match generation on an array-by-array basis. Best-match selection can also be synchronized to this array output rate. PMF and Prazdny, on the other hand, require all, or a good portion, of the candidate match arrays before local support computation can proceed. These two algorithms use voting that extends to candidate matches at all disparities, which makes the local neighborhoods very large. Although certain ingenious simplifications are employed to reduce the computational burden, significantly more memory is nevertheless required in these two algorithms. Due to memory access bottlenecks and nonsynchronous computation between different algorithm steps, the throughput and latency of these two algorithms are expected to be much lower than MPD. For these reasons MPD showed best promise for a fast, compact (analog) VLSI implementation.

The improved speed potential of MPD is not achieved without some penalty, however; it can be argued [12], [14] that MPD sacrifices accuracy to increase throughput. This issue has been investigated in detail in [14] with the result that if the scene distance is a few times larger than the interocular separation, disparity becomes a fairly smooth function over the image and then MPD performance is expected to be comparable to PMF and Prazdny. (This is really only true for "opaque" scene objects—no transparent or "partially transparent" structures in the foreground such as a picket fence or a fine grating.) Recall that since the image planes were initially assumed to be coplanar (independent of algorithm choice), distance to the scene was constrained to be a few times larger than the interocular separation. For the coplanar image geometry, therefore, MPD is generally equivalent to PMF and Prazdny in performance and thus MPD appears to be a good overall compromise of speed and accuracy for the present case.

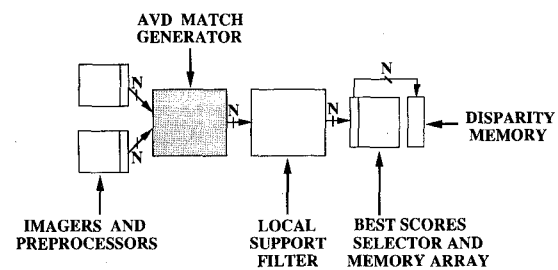


Fig. 3. CCD/CMOS stereo vision system.

A detailed simulation study of the MPD algorithm was performed to investigate VLSI performance issues [14], [15], particularly in relation to analog processing. The results suggested that many of the traditional disadvantages of analog VLSI (limited accuracy, inflexibility, and lack of storage capacity) can be avoided with this algorithm. This was demonstrated with three particular simulation results: 1) the computational accuracy requirement in each of the four algorithm steps is less than an 8-b equivalent level without any observable sacrifice in system accuracy; 2) the scene dependent programmability requirements of the algorithm are modest; and 3) the memory requirement is small (due to synchronous candidate match data, local support, and best-match selection processes). A typical simulated disparity map output of the MPD stereo system is shown in Fig. 2 together with the  $256 \times 256$  pixel input image pair. Disparity is encoded as brightness (brighter is closer). All data were restricted to 8-b accuracy during simulation. Note that the larger object (teddy) is detected well while the thin object (toy crane arm) is partially lost. A possible MPD stereo system implementation is discussed next.

### III. CCD/CMOS STEREO SYSTEM ARCHITECTURE

A full CCD/CMOS implementation of the MPD algorithm, shown in Fig. 3, consists of seven modules: two imagers (left and right views), two image filters, a match data generator (also called the shift-and-compare correlator), a local support filter, and a best-match selector. The system blocks correspond closely with the four algorithm steps as can be seen by comparing Fig. 3 and the algorithm outlined in the previous section.

A striking feature of the MPD algorithm is that it, like many (early) vision algorithms, utilizes only local interactions between neighboring processor cells. Furthermore, the local

interactions are typically repeated simultaneously over large sections of the images. Such a computation pattern can readily take advantage of parallel processor arrays, using one processor per pixel. To maximize the power of the parallel processor, therefore, special attention is paid to communication pathways in the system. In all modules of the MPD system, at most three types of communication exist: 1) processor-to-processor, 2) processor-to-local memory, and 3) I/O's between different system modules. The processor-processor interactions occur along a simple North-East-West-South (NEWS) grid, which can be realized efficiently in hardware. Processor-local memory operations are equally simple, since at any step in the algorithm a given processor only needs to keep one (analog) data value in its own memory. It will be seen that a CCD potential well can readily be used as a local analog memory cell. The heaviest data transfer burden in this system thus falls on intermodule I/O operations, which are array-by-array synchronous. To maximize the data communication speed, column parallel I/O structures are used.

CCD/CMOS technology appeared to be best suited for analog implementation of the MPD stereo system, mainly because CCD's can be used as small and accurate analog memory cells and shift registers. The shift registers need to implement a programmable shift-and-compare operation (see Section IV), and cannot be obtained as easily and compactly from other typical analog technologies. Moreover, the availability of CMOS devices from the same process allowed efficient signal handling outside the CCD channel. All necessary operations of the MPD algorithm (shifting, summation, subtraction, division by a constant, and comparison) can be performed efficiently with CCD/CMOS circuitry. When the system clock speed is sufficiently fast (in the megahertz range), dark current effects can be eliminated and CCD memory cells do not need a refresh operation. The following paragraphs describe system blocks that could be used to implement the CCD/CMOS stereo system shown in Fig. 3. The preprocessor (step 1) and local support (step 3) modules have already been reported in previous work [3], [16]–[18]. The match generator module (step 2, shaded in Fig. 3) and the best-match selector (step 4), on the other hand, have not yet been demonstrated. The various system blocks described next.

*Algorithm step #1:* Image preprocessing is done with spatial bandpass filters that emphasize brightness edges in the images and suppress both spatial high-frequency noise and low-frequency terms. The spatial bandpass filters can be implemented using a cascade of a low-pass binomial convolver and a Laplacian high-pass filter, yielding "Laplacian of Binomial" or LoB filter. A binomial filter of adjustable spatial extent is achieved with repeated applications of the kernel

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (1)$$

while a Laplacian operation requires one application of the following kernel at each pixel:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \quad (2)$$

Both the Laplacian and the binomial sections of the filter can be constructed using nearest-neighbor-only communication between pixels. A CCD imager and its LoB preprocessor can be integrated on the same chip without loss of imager area, as demonstrated in [16]. This implementation, however, fixes the extent of the binomial filter in hardware. Other CCD implementations that have programmable filter extents of the binomial convolver have been reported in [17], [18]. Image preprocessing for the stereo system can be done with either type of binomial filter.

*Algorithm steps #2 and #3:* The candidate match generator (step #2) and the local support filter (step #3) execute the largest number of operations in the algorithm, thus directly dictating the speed of the whole system. In Section II the local support filter was explained to have a simple spatial low-pass characteristic. A binomial convolver with a programmable filter extent is a suitable choice. While this filter has already been fabricated, a CCD circuit capable of functioning as a match generator (shift-and-compare correlator) had not been reported. For these reasons, it was selected for design and fabrication. Candidate matches in this design are generated with an AVD array processor that compares each pixel in one preprocessed  $N \times N$  (e.g.,  $40 \times 40$ ) image with a range of pixels in the other preprocessed  $N \times N$  image. Comparison over a range of pixels is achieved by repeated compare-shift operations (one preprocessed image is shifted while the other is held still). Each comparison step produces one  $N \times N$  array of AVD results that are processed by the local support filter to produce local support scores. The AVD module is described in detail in Section IV.

*Algorithm step #4:* In the last algorithm stage, each  $N \times N$  array of local support scores is compared with current values in the  $N \times N$  best scores array (an analog memory). This comparison is done pixel-wise in a column-parallel fashion. For each local support score array produced, a digital counter keeps track of the current shift (= disparity) value. Each time a better score is found, both the best score array and the disparity array (which contains the best current disparity values in a digital memory) are updated at that location. The support score comparison needs a column of  $N$  comparators as well as  $N$  write signals. Each comparator should be accurate to about 8 b. For a 2-V input range of the comparator this implies an offset of less than 8 mV. This is readily available with a CMOS comparator. With each comparison phase the scores in the best scores memory array need to be recycled either by physically arranging the CCD memory to be rotating or by exploiting the charge-by-wire transmission idea proposed in [19].

Note that the output of the stereo algorithm is fully contained in the digital  $N \times N$  disparity memory array. The analog results (local support scores in the best scores array), on the other hand, only provide a write instruction for the digital memory, but are not needed as outputs. Consequently, this stereo system does not require any analog-to-digital converters. In the present system, the AVD processor has been implemented in a CCD/CMOS process, while other blocks are emulated by a computer, as will be discussed in Section V.

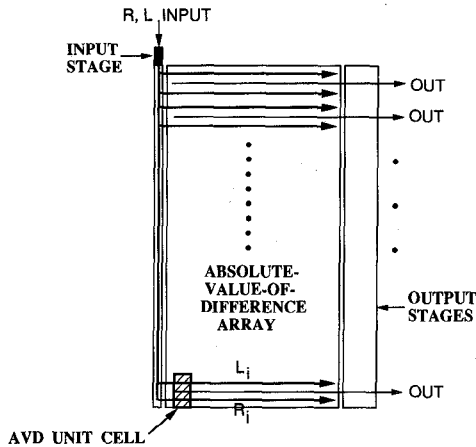


Fig. 4. Basic data paths in the  $40 \times 40$ -pixel AVD match generator.

#### IV. CCD/CMOS AVD PROCESSOR

A  $40 \times 40$ -pixel version of the match generator module was implemented as an AVD array in a  $2\text{-}\mu\text{m}$  CCD/CMOS process (see Fig. 4). The total chip area is  $7.4 \times 8.7 \text{ mm}^2$  and its on-chip power dissipation is about  $450 \text{ mW}$  ( $f_{clk} = 2 \text{ MHz}$ ). The array can perform an AVD operation between two (preprocessed)  $N \times N$  images a programmable number of times. It is operated by clocking CCD charge packets through the following signal processing stages:

- 1) CCD input stage which performs voltage-to-charge conversion (inputs to the AVD array are in the voltage domain).
- 2) CCD shift registers for alignment of left and right image rows into a paired-row formation as shown in Fig. 4. In this manner, corresponding left and right image rows (from Fig. 1) end up in adjacent shift registers.
- 3) Floating-gate output stage for nondestructive sensing of signal charge in the left and right rows (charge-to-voltage conversion).
- 4) AVD input stage (voltage-to-charge conversion). Floating-gate output voltages are used to compute individual pixel-level AVD values. The AVD results are stored as charge packets in a third set of CCD shift registers (labeled "OUT" in Fig. 4), which are subsequently used to clock out the results.
- 5) Floating diffusion output for destructive readout of AVD results (final charge-to-voltage conversion).

After this cycle is completed, one set of data (e.g., the left registers) is shifted by one pixel, keeping the other set still. The AVD operation is then repeated from stage 2, and the next set of match data is generated. This process continues until a sufficiently large number of shift-and-compare cycles have been completed. By controlling the number of shifts, a user-programmable amount of candidate match arrays can be generated. Each of the stages is described in more detail next.

##### A. CCD Input Stage

The "fill-and-spill" input structure consists of a total of five buried-channel CCD (BCCD) gates. The input is shown in cross section in Fig. 5(a), where the channel potential is plotted

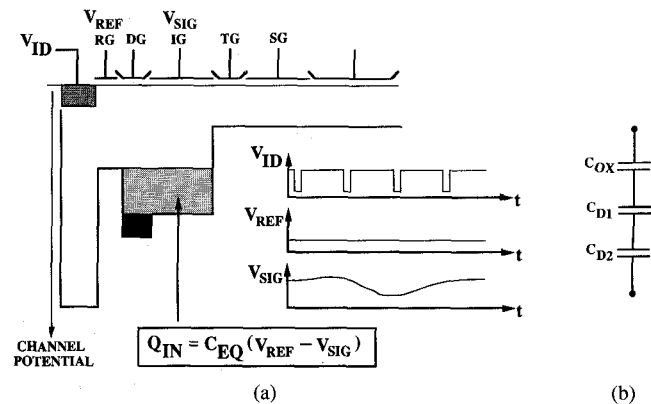


Fig. 5. (a) "Fill-and-spill" BCCD input structure in cross section. (b) Equivalent circuit for a single BCCD gate.

in the vertical direction versus distance on the horizontal axis. Two of the gates ( $RG$  and  $IG$ ) are used for actual charge metering, another two ( $TG$  and  $SG$ ) are used for charge transfer into the CCD channel, and one gate acts as a "dummy" ( $DG$ ). The input structure shown differs from typical "textbook" input circuits [20]–[22] by the addition of the dummy gate  $DG$  as well as the transfer-storage gate pair  $TG$ – $SG$  [23]. The other two gates ( $RG, IG$ ) operate like textbook fill-and-spill inputs: while a signal  $V_{SIG}$  is held at the input gate  $IG$  and a reference voltage  $V_{REF}$  is applied to the reference gate  $RG$ , the input diffusion ( $ID$ ) voltage  $V_{ID}$  is pulsed low, filling the channel under  $RG, DG$ , and  $IG$  with electrons (see timing diagram in Fig. 5(a)). When  $V_{ID}$  is brought high again a moment later, some charge spills from the channel region back to the input diffusion while an amount of charge determined by the potential barriers under  $RG$  and  $IG$  is left behind. The input charge is found approximately using

$$Q_{SIG} \approx C_{EQ}(V_{SIG} - V_{RG}) \quad (3)$$

where  $C_{EQ}$  is found from the equivalent model of a BCCD gate shown in Fig. 5(b):

$$C_{EQ} = \frac{C_{OX}C_{D1}}{C_{OX} + C_{D1}} \quad (4)$$

Note that both  $C_{D1}$  and  $C_{D2}$  model capacitances involving depleted semiconductors, and thus both are nonlinear (only  $C_{D1}$  appears in  $C_{EQ}$ , however).

The function of  $DG$  is to allow charge input to be measured using gates made with the same polysilicon level, thus eliminating offsets due to differences in oxide thickness between poly 1 and poly 2 (note that the excess charge contained under  $DG$  does not contribute to input charge). The  $TG$ – $SG$  pair, on the other hand, operates to "scoop" charge from the input gates without fear of backspilling upon subsequent charge transfer to the actual CCD channel. This is accomplished by keeping  $V_{TG}$  smaller than  $V_{SG}$  when  $V_{SG}$  goes low. This design uses BCCD gates at the input because the CCD process supported by MOSIS does not provide surface channel devices. The input nonlinearity due to  $C_{EQ}$  was calculated to be less than 10%, which does not affect performance in the MPD stereo system.

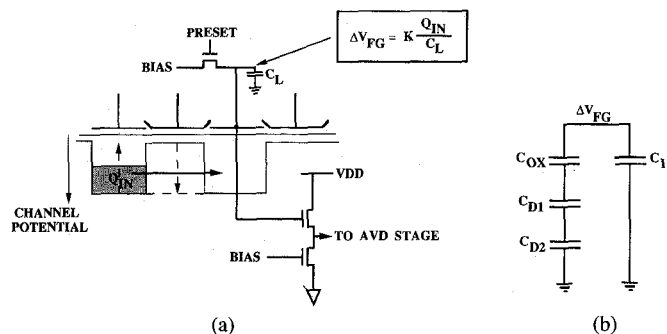


Fig. 6. (a) FG output circuit in cross section. (b) Equivalent circuit of FG node.

### B. CCD Shift Registers

As mentioned above, the CCD shift register clocks are controlled such that the input charge packets from the left and right images are properly lined up in their respective rows. The charge transfer efficiency of the BCCD devices was found experimentally to be 0.9992 (measured using 488 gates of 14  $\mu\text{m}$  length each). The longest transfer path in the AVD array is 800 gates, resulting in a maximum of 6% charge transfer loss. Due to the differential nature of the AVD cell (left and right charge packets in neighboring rows travel about the same distance), the charge transfer efficiency is sufficiently good for proper operation of the array.

### C. Floating-Gate Output Stage

The floating-gate (FG) sense technique [13] is used to measure charge in the left and right image rows, converting channel charge into a voltage output. A floating gate rather than a floating diffusion output is used because the sensing must be carried out nondestructively (the same charge buckets in the left and right rows are used over and again for different comparison cycles). The FG circuit shown in Fig. 6(a) senses charge in the CCD channel capacitively. The channel potential of the sensing device changes when a charge packet is introduced into its potential well. If the gate terminal of this device is floating at this time (initially precharged to a high potential), the gate voltage will also change due to capacitive coupling through the semiconductor and the oxide. An equivalent circuit for the FG circuit is shown in Fig. 6(b). It should be noted that in general the maximum charge that can be introduced into the FG potential well may need to be smaller than the maximum charge capacity of the CCD device. Qualitatively, this occurs because the change in the gate potential of the FG node causes a further reduction in the channel potential, which was already reduced by the signal charge packet. In normal shift-register operation the gates are always tied to voltage sources, in which case the floating-gate effect causing gate voltage reduction is not observed.

The BCCD FG voltage change can be found approximately by applying the linearized capacitor model shown in Fig. 6(b). This yields the result

$$\Delta V_{FG} \approx \frac{Q_{SIG}}{C_L} \times \frac{1/C_{D2}}{1/C_{OX} + 1/C_L + 1/C_{D1} + 1/C_{D2}} \quad (5)$$

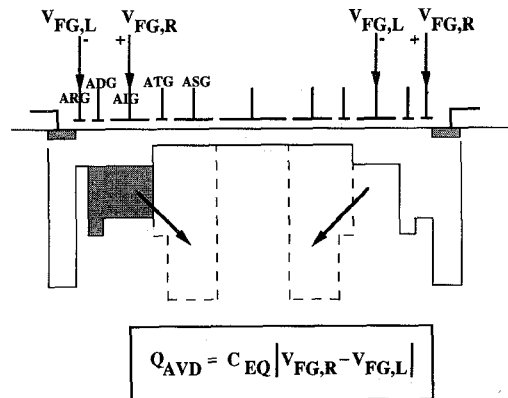


Fig. 7. Cross section of AVD input stage (for the case shown,  $V_{FG,R} > V_{FG,L}$ ).

where  $Q_{SIG}$  is the measured signal charge. Note again that  $C_{D1}$  and  $C_{D2}$  are nonlinear capacitors. In most BCCD structures  $C_{OX} \gg C_{D2}$  and also  $C_{D1} > C_{D2}$  with a comfortable margin (about 5). In such a case increasing the load capacitance  $C_L$  to about  $C_{OX}$  has the beneficial effect that the charge capacity of the floating-gate node is increased and the transfer function is made more linear. Further increases in  $C_L$  are not as beneficial, however, since the dynamic range of the FG node is compromised without significant improvement in linearity. The present design has a parasitic load capacitance of about  $C_{OX}$  due to adjacent CCD gate overlap as well as the source-follower input capacitance, and thus did not require an additional capacitor at the FG node. Finally, the source-follower (used to buffer the FG output, see Fig. 6(a)) gain was measured to be about 0.95.

### D. AVD Stage

The AVD stage shown in Fig. 7 consists of two cross-coupled fill-and-spill input circuits similar to those discussed earlier. In the array the input and reference gates of the AVD stage (AIG and ARG) are driven by floating-gate circuit outputs of the left and right image rows. Since the AVD inputs are cross-coupled, one fill-and-spill stage receives an input charge proportional to the absolute value of difference between the left and right pixel values, while the other fill-and-spill stage is empty. Since it is not known which of the input stages receives charge and which is empty, a subsequent addition of the two charge packets completes the AVD operation. The addition is conveniently performed in the shift register (OUT) that is also used to shift out the AVD results.

A single AVD unit cell plan view is shown in Fig. 8, which contains a single shift-register stage of a left (top) and right (bottom) image row (see also Fig. 4). FG nodes are shown as well (shaded), including the associated source followers. The middle structure contains the two AVD fill-and-spill stages as well as one stage of the shift register used to store and shift out the AVD results. Control lines for various CCD gates are also indicated in the figure.

### E. Floating Diffusion Output Stage

The floating diffusion (FD) circuit shown in Fig. 9 is a typical termination structure for a CCD channel and is

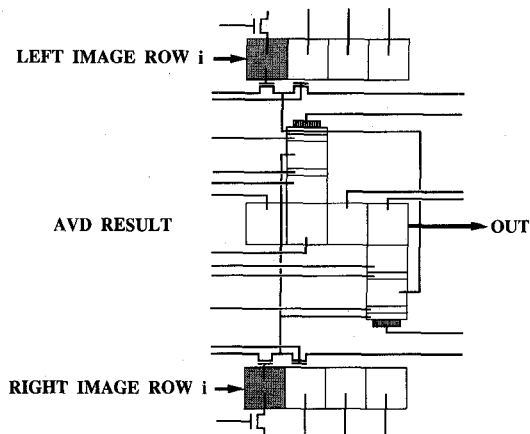


Fig. 8. AVD unit cell plan view.

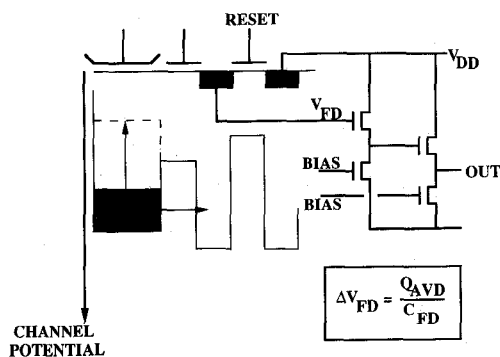


Fig. 9. Cross section of FD output circuit.

analogous to the charge input circuit [20]–[22]. The output stage used in this design is fairly standard, the use of a buried-channel MOSFET (BMOS) as the reset gate being the most notable feature. The BMOS device allows sharing of the floating diffusion between the CCD channel and the reset gate, which reduces the output capacitance as compared to an enhancement MOSFET as the reset gate. A small output capacitance is desirable to achieve sufficiently high signal swings since the FD voltage change is given by

$$\Delta V_{FD} \approx \frac{Q_{SIG}}{C_{FD}}. \quad (6)$$

The FD is buffered with a standard two-stage source-follower circuit which has a gain of about 0.89. (Nonlinearities introduced by the junction capacitor  $C_{FD}$  and the source-follower back-gate effect were unimportant in the stereo algorithm.)

## V. AVD PROCESSOR TEST RESULTS

The overall signal flow for the full AVD array is depicted in Fig. 10. The gain factors shown for each stage are (linearized) average values from measurements of the unit cells and correspond to (3)–(6). The overall gain from the input CCD's to the floating diffusion outputs is found to be about 0.8 (i.e., a voltage difference of 1 V between a left image pixel and a right image pixel gives an absolute value of difference output

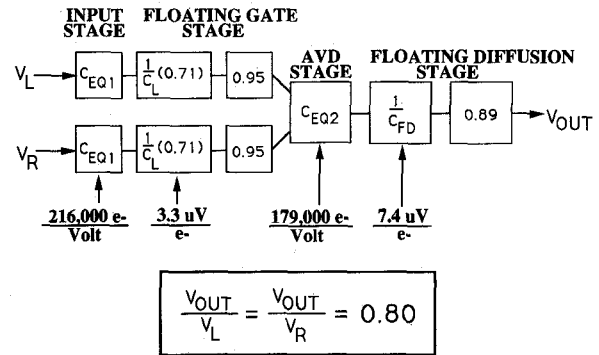


Fig. 10. Signal flow summary for AVD array.

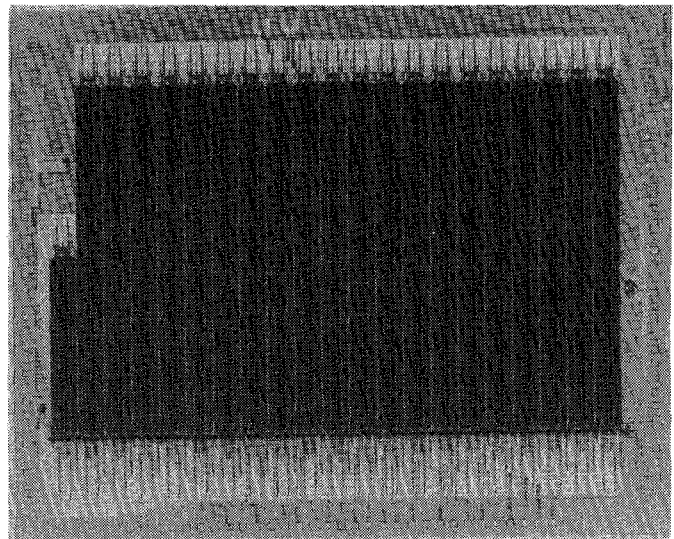


Fig. 11. Die photomicrograph.

of 0.8 V). The photomicrograph of the chip containing the AVD processor and several test structures is shown in Fig. 11.

The AVD array functionality as a match generator module in a stereo algorithm environment was tested by interfacing it with a computer that performed the other three algorithm steps [14]. These results were compared against a full computer simulation of the algorithm. A special high-speed test board was constructed to thoroughly exercise the AVD chip. Both computer-generated and real-scene input images were used; the real-scene input pair is shown in Fig. 12. Results of the real-scene test are illustrated in Fig. 13 where both the test system output (which uses the AVD chip) and a full computer simulation output (which does not use the AVD chip) are shown. The stereo output is a disparity map with distance to the scene encoded as brightness, which gives a contour-like impression (bright = near, dark = far). The two outputs are seen to correspond closely, which confirms the overall functionality of the AVD chip in the stereo system. At a 2-MHz clock rate for the CCD processor, approximately 160 input frame pairs can be processed per second in the current implementation, which uses serial data input for test purposes. A full stereo system implementation would avoid the input bottleneck by utilizing column parallel inputs. This should increase the processing speed of the AVD processor up to



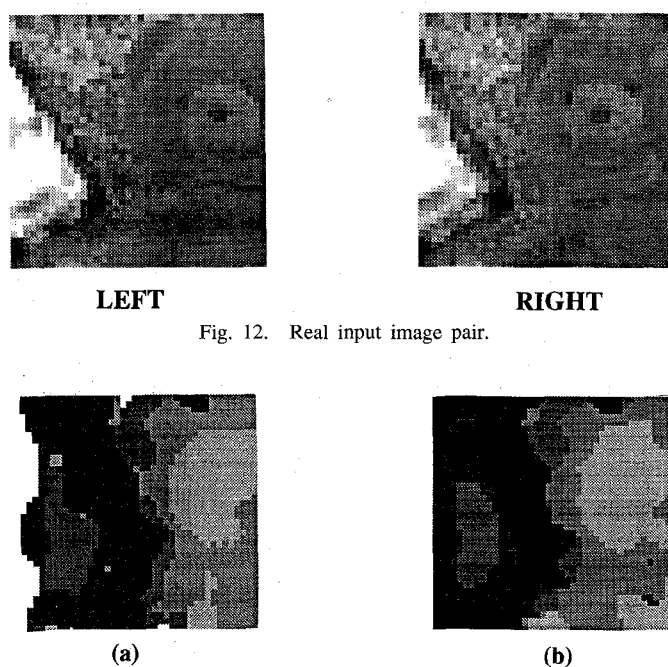


Fig. 12. Real input image pair.

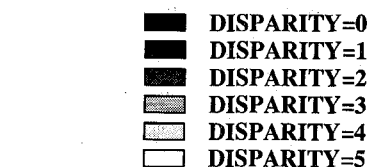


Fig. 13. (a) System output with AVD chip. (b) Simulated system output.

3200 frame pairs per second, allowing a total stereo system speed of about 880 frame pairs per second for a disparity (shift) range of 10 pixels and a local support (binomial) filter extent of 15 pixels.

Note, however, that the test system output (using the AVD chip) does not exactly match the fully simulated output at all pixel locations. The discrepancies were found to be due to mismatches between neighboring AVD unit cells. The mismatches were measured by operating the AVD chip with constant (but unequal) value inputs. The  $N \times N$  output values were observed, which in the absence of mismatches would have been equal (ignoring noise). The measured errors between individual outputs averaged about  $\pm 8\%$ , which was consistent with a predicted value of about  $\pm 7\%$ . The main contributors to this error were found to be a) cell-to-cell AVD stage input capacitance mismatch manifested as a gain mismatch (equation (3)) ( $\Delta C_{EQ}/C_{EQ}$  assumed to be 1%, accounting for about 65% of the total error); b) cell-to-cell FG readout stage gain mismatch (equation (5)) ( $\Delta C_L/C_L = 1\%$ , for 18% of total error budget); and c) cell-to-cell FG source-follower offset ( $\Delta[W/L]/[W/L] = 1\%$ ,  $\Delta V_T = 5$  mV, for 11% of total error). The remaining 6% of the error budget was due to FG preset switch charge injection mismatch and thermal noise.

An additional effect of the CCD input unrelated to the offsets described above was observed. The CCD input structure is known to suffer from "dead-zone" nonlinearity due to thermionic back-emission of charge from the input potential well to the input diffusion [18]. This phenomenon is best observed when the difference between the input and reference

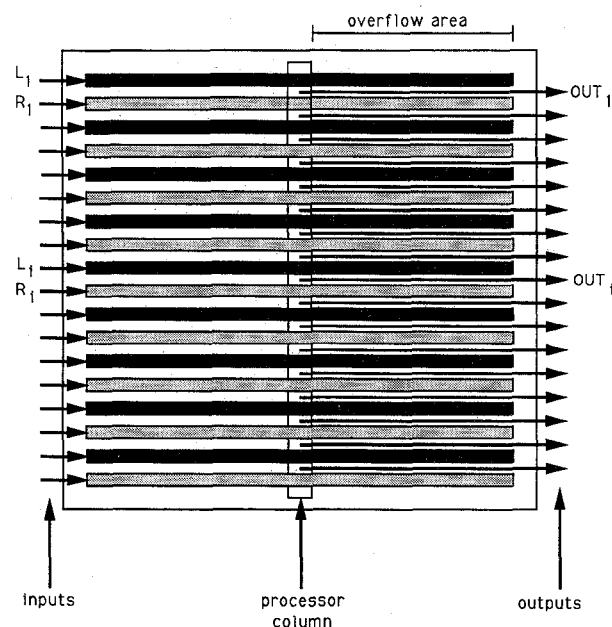


Fig. 14. New AVD match generator chip architecture.

voltages is quite small (see (3)). Higher energy electrons in a full input well do not see a large enough potential barrier and are lost through random thermal motion back to the input diffusion (typically several  $kT/q$ 's worth of electrons are lost). Note that virtually no charge can be emitted in the reverse direction since electrons at the input diffusion see a very large potential barrier, making this phenomenon unidirectional. Measurements indicated that any voltage difference less than 0.3 V (on the order of  $10 kT/q$ ) between the input and reference gates produced a nearly empty charge packet. The input degradation is frequency dependent: the amount of charge lost is reduced if the input charge is clocked quickly into the channel. Note that the impact of this effect on the system can be reduced by introducing a deliberate offset to the charge packets at the chip inputs. The same approach is unfortunately not possible in the AVD unit cell due to the cross-coupling of input gates which precludes the deliberate use of a dc offset.

Despite the observed offsets and the input nonlinearity, the overall performance of the AVD chip in the stereo system is encouraging. Extensive simulations confirm that the MPD algorithm is quite forgiving of these errors. However, a possible way to improve accuracy is illustrated in Fig. 14, where the  $N \times N$  array of AVD cells is replaced by a column of  $N$  cells. This should help in reducing cell-to-cell offsets within the same row since all operations are performed with the same processor. The row-to-row mismatch can also be improved relative to the present AVD cells by redesigning the AVD stage (where  $\Delta C_{EQ}/C_{EQ}$  was the main culprit) in CMOS outside the CCD channel. This can also eliminate the "thermionic effect" nonlinearity in the AVD stage. Furthermore, the AVD operation can be replaced with another metric, such as the square-of-difference, which may be more attractive for CMOS implementation. An added benefit of the new design is a reduced total array size because of a smaller amount of



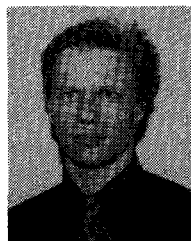
processors. For the CCD gate sizes of the current design, up to  $100 \times 100$  images could be processed with a  $1\text{-cm}^2$  chip area. A disadvantage of the proposed design is the need for charge recycling, implemented with an "overflow" area shown in Fig. 14. The overflow problem could possibly be avoided with the charge-by-wire scheme reported recently [19].

#### ACKNOWLEDGMENT

MOSIS provided the fabrication. The authors thank Prof. J. L. Wyatt and Prof. C. G. Sodini for their helpful suggestions.

#### REFERENCES

- [1] J. Wyatt *et al.*, "Analog VLSI systems for image acquisition and fast early vision processing," *Int. J. Comput. Vision*, vol. 8, no. 3, pp. 217-230, Sept. 1992.
- [2] E. R. Fossum, "Charge-coupled computing for focal plane image pre-processing," *Opt. Eng.*, vol. 26, no. 9, pp. 916-922, Sept. 1987.
- [3] W. Yang and A. M. Chiang, "VLSI processor architectures for computer vision," presented at the DARPA Image Understanding Workshop, Palo Alto, CA, May 1989.
- [4] C. A. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.
- [5] M. Drumheller and T. Poggio, "On parallel stereo," in *Proc. IEEE Int. Conf. Robotics Autom.*, vol. 3 (San Francisco, CA), 1986, pp. 1439-1448.
- [6] *Digital Signal Processing Data Book*, LSI Logic Corp., Milpitas, CA, 1990.
- [7] T. Poggio, "Vision by man and machine," *Scientific American*, vol. 250, no. 4, pp. 106-116, Apr. 1984.
- [8] B. K. P. Horn, *Robot Vision*. Cambridge, MA: MIT Press, ch. 13.
- [9] B. K. P. Horn, "Height and gradient from shading," *Int. J. Comp. Vision*, vol. 5, no. 1, pp. 37-75, 1990.
- [10] K. Prazdny, "Egomotion and relative depth map from optical flow," *Biol. Cybern.*, vol. 36, no. 2, pp. 87-102, 1980.
- [11] S. T. Barnard, "Stochastic stereo matching over scale," *Int. J. Comp. Vision*, vol. 3, pp. 17-32, 1989.
- [12] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby, "PMF: A stereo correspondence algorithm using a disparity gradient limit," *Perception*, vol. 14, pp. 449-470, 1985.
- [13] K. Prazdny, "Detection of binocular disparities," *Biol. Cybern.*, vol. 52, pp. 93-99, 1985.
- [14] J. M. Hakkarainen, Ph.D. dissertation, Mass. Inst. of Technol., Cambridge, 1992.
- [15] J. M. Hakkarainen *et al.*, "Interaction of algorithm and implementation for analog VLSI stereo vision," in *Proc. SPIE Conf. 1473 Visual Inf. Proc.: From Neurons to Chips* (Orlando, FL), Apr. 1991, pp. 173-184.
- [16] W. Yang, "Analog CCD processors for image filtering," in *Proc. SPIE Conf. 1473 Visual Inf. Proc.: From Neurons to Chips* (Orlando, FL), Apr. 1991, pp. 114-127.
- [17] J. P. Sage and A. Lattes, "A high-speed analog two-dimensional Gaussian image convolver," in *Tech. Dig. Topical Meeting Machine Vision* (Incline Village, NV), Mar. 1985, p. FD5-1.
- [18] C. L. Keast, Ph.D. dissertation, Mass. Inst. of Technol., Cambridge, 1992.
- [19] E. R. Fossum, "Wire transfer of charge packets using a CCD-BBD structure for charge-domain signal processing," *IEEE Trans. Electron Devices*, vol. 38, no. 2, pp. 291-298, Feb. 1991.
- [20] D. K. Schroder, *Advanced MOS Devices: Modular Series on Solid State Devices*, G. W. Neudeck and R. F. Pierret, Eds. Reading, MA: Addison-Wesley, 1987.
- [21] C. H. Sequin and M. F. Tompsett, *Charge Transfer Devices: Advances in Electronics and Electron Physics*, suppl. 8. New York: Academic, 1975.
- [22] C. K. Kim, "The physics of charge coupled devices," in *Charge-Coupled Devices and Systems*, M. J. Howes and D. V. Morgan, Eds. New York: Wiley, 1979.
- [23] A. M. Chiang, personal communication, MIT Lincoln Laboratories, Bedford, MA, 1990.



**J. Mikko Hakkarainen** (M'93) was born in Kaajaani, Finland, on March 6, 1962. He received the S.B. and S.M. degrees in 1988 and the Ph.D. degree in 1992 in electrical engineering from the Massachusetts Institute of Technology, Cambridge. His Ph.D. dissertation was titled "A Real-Time Stereo Vision System in CCD/CMOS Technology."

He worked at Analog Devices in Wilmington, MA, from 1984 to 1991 first as a co-op student and later as an engineer, primarily engaged in silicon IC magnetic field sensor design and phase-locked loop modeling. In 1991 he joined the research staff at General Electric Corporate Research and Development, Schenectady, NY, where he designs analog and mixed mode circuits for sensor, low-power, and high-temperature applications.



**Hae-Seung Lee** (M'85-SM'92) was born in Seoul, South Korea, in 1955. He received the B.S. degree in electrical engineering from Seoul National University, Seoul, Korea, in 1978. In 1980 he received the M.S. degree in electrical engineering with emphasis on feedback control systems from the same university. He received the Ph.D. degree in electrical engineering from the University of California, Berkeley, in 1984, where he developed self-calibration techniques for A/D converters.

In 1980 he was employed as Technical Staff in the Department of Mechanical Engineering at the Korean Institute of Science and Technology, Seoul, Korea, where he was involved in the development of alternative energy sources, and modeling of magnetic bearings. In 1984 he joined the faculty in the Department of Electrical Engineering and Computer Science at Massachusetts Institute of Technology, Cambridge, where he is now a Professor. Since 1985 he has acted as a consultant to Analog Devices Semiconductor, Wilmington, MA, and MIT Lincoln Laboratories, Lexington, MA. His research interests are in the areas of integrated circuits including data converters, signal processing and communication circuits, and solid-state sensors.

Prof. Lee is a recipient of the 1988 Presidential Young Investigators' Award. He has been an Associate Editor of the IEEE Journal of Solid-State Circuits since 1992.