Digital image processing systems based on functional-oriented processors with a homogeneous structure

N A Lookin

Institute of Engineering Science, Ural Branch of RAS, 34 Komsomol street, Yekaterinburg, 620049

E-mail: nicklookin@mail.ru

Abstract. Real-time digital image processing as part of mobile systems requires solving the problems of development high-performance specialized processors for intraframe image processing. The graphs of intraframe processing algorithms are analysed and their parameters are evaluated, primarily the degree of parallelism. The architectures of functional-oriented processors based on homogeneous computing environments for ultrafast intraframe processing are described. The results of research and development of microelectronic implementation of such processors are presented.

1. Introduction

Digital image processing is one of the most relevant and intensive computational procedures in real-time systems. Sometimes a real-time mode may make general-purpose processor platforms unacceptable [1]. In some cases, it is difficult to apply even such advanced solutions as multiprocessor GPU architectures [2]. An increase in the frame rate of optical sensors and the number of output channels for picking up pixels from the array of sensitive elements may also require a significant change in the architecture of video processors for embedded image processing systems. This paper discusses some of the key problems in the development of mobile image processing systems, as well as describes the architecture and microelectronic implementation of specific video processor with massively parallel data processing.

2. Features of real-time digital image processing algorithms

The one of the main features of image processing algorithm for most mobile real-time systems is two-stage data processing.

Stage 1. Preliminary or intraframe processing, the purpose of which is to increase the reliability of the images. At this stage, algorithms for correcting geometric distortions, increasing contrast, averaging, smoothing, filtering, segmentation are implemented [3]. The special feature of intraframe processing algorithms is a significant number of independent data streams and a high degree of internal parallelism of computations. For example, for a CCD matrix with a size of 2Kx2K pixels, the maximum possible parallelism is 4·10^6 independent data streams.

Stage 2. Final (interframe) processing, the purpose of which is to identify the parameters of external objects and, ultimately, to recognize them. At this stage, algorithms for computation image attributes, formal invariants, determining the position and parameters of motion of objects in various coordinate systems, and constructing models of the problem environment are implemented. An example of the interframe image processing is the problem of the motion parameters estimation of an object external...
to the video system (a well-known problem named as "Motion Estimation") [4]. The features of such algorithms are lower degree of parallelism, a significant number of data iterations, and high connectivity of algorithmic graphs. The peculiarities of data processing algorithms, which are important for implementation in computers, are closely related to the structures of their information graphs $G$ [5]. The main parameters of algorithmic graphs are as follows:

a). Degree of Parallelism (DOP) [6]. In order to correctly estimate the DOP, the algorithmic graphs must be transformed to multi-level parallel form (MLPF) [7]. After that current value of parameter $DOP_h$ is estimated, where $h$ is the level number. It can be shown that $\forall G, \exists m, H \in \Gamma; DOP(h_m) = \text{Max}(DOP(h)), h = 0, 1, ..., H$.

b). Width (diameter) $D$ of the MLPF [5], which can be used as a measure of the hardware complexity of computations ($L_h$) [8].

c). The height $H$ of the MLPF [5], which is determined by the number of levels. The parameter $H$ can be used as a measure of the time complexity of computations ($L_t$) [8].

For most intraframe image processing algorithms $D = N1 \cdot N2$, where $N1$, $N2$ is the number of pixels of the sensor’s photosensitive matrix along the $i$ and $j$ coordinates. For example, if $N1 = N2 = 2K$, then $D = 4 \cdot 10^6$. It means the possibility of parallel processing of four million data streams. The main reason is the functional independence of the pixels of the sensor matrix: $p_{i,j} \neq f(p_{i\pm1,j\pm1})$ where $p_{i,j}$ is the amplitude of the brightness of the reflected signal in the cell with the coordinates $i, j$. Thus, intraframe image processing algorithms have the maximum parallelism.

Example. Consider the simplest algorithm for two-dimensional image filtering for window size $3 \times 3$:

$$
(p_f)_{i,j} = \frac{1}{9} \sum_{i-t}^{i+t} \sum_{j-t}^{j+t} p_{i,j} \cdot c_{i,j},
$$

where $p_{i,j}$ is the pixel of the original image; $c_{i,j}$ - filter kernel element; $(p_f)_{i,j}$ - pixel value of the filtered image; $i = 0, 1, ..., (N1-1)$; $j = 0, 1, ..., (N2-1)$. The MLPF of a graph for filtering of an image size $3 \times 3$ is shown in Figure 1.

![Figure 1. The algorithmic graph MLPF of the 2D-filtering algorithm for an image size $3 \times 3$.](image)

Analysis of the algorithm. The main graph parameters: $D = 9, H = 6$. The minimum time for one-time filtering implementation is $6t$ ($t$ is the time of an arithmetic operation) on 9 processors, so hardware computational complexity $L_h = 9$. Current $DOP$ values for each level of the graph are also shown on Fig. 1. Algorithmic complexity for this graph $L_h = 16$. Autonomous time complexity of calculations $L_{Aut}$ corresponds to the one-time data processing, pipeline time complexity $L_{Conv}$ - to continuous processing of data streams entering the system input. For one-time filtering the time interval from the arrival of the arrays $\{p_{i,j}\}$ and $\{c_{i,j}\}$ to the output of the parameter $(p_f)_{i,j}$ is $6t$, i.e. $L_{Aut} = 6t$. For continuous filtering this interval decreases to $1t$, i.e. $L_{Conv} = t$.

The parameters of the algorithmic graph of 2D-filtering of images size $3 \times 3$ and estimates of the computational complexity are given in Table. 1.
Table 1. Graph parameters for 2D image filtering algorithm and computational complexity estimation.

| MLPF parameters | $D$ | $H$ | $L_{ina}$ | $L_{conv}$ | $L_{i}$ | $L_{a}$ |
|------------------|-----|-----|-----------|------------|--------|--------|
| Image size 3x3   | 9   | 6   | 6         | 1          | 9      | 16     |
| Image size N1xN2 | $N1 \cdot N2$ | 6   | 6         | 1          | $N1 \cdot N2$ | 16 $N1 \cdot N2$ |

A 2D filtering algorithm for a full-size image ($N1 \times N2$ pixels) requires the implementation of approximately $N1 \times N2$ such filters. The graph of the algorithm (1) in the form of the maximum LPF is shown in Figure 2.

Figure 2. MLPF of the graph for 2D filtering images algorithm (image size $N1 \times N2$, filter kernel 3x3).

Thus, the MLPF of the 2D filtering algorithm graph for full-size images has the maximum possible width $D \approx N1 \cdot N2$, the graph height is minimal: $H = 6$. Algorithmic graph parameters are given in Table. 1. With maximum parallelization, the image size does not affect the time complexity of the calculations.

3. Processor architectures and real-time image processing

As mentioned above, the base of almost all algorithms for intraframe real-time image processing is 2D filtering, which makes creates the possibility of maximum possible parallelism. Mathematical models of algorithms for the intraframe image processing can be represented in the form of ensembles of Moore automata [9].

Interframe image processing algorithms are iterative, their graphs have a significant number of data cycles, therefore, for these algorithms, the time complexity of computations depends on the number of loop repetitions. This corresponds to model of the Mealy automata [9]. The algorithms of this stage of image processing have significantly less potential parallelism and have significantly lower algorithmic complexity compared to intraframe processing.

A rational solution for data processing in mobile video systems is a hybrid processor architecture that considers the features of the image processing algorithms noted above. The functional modules of this architecture are:

- In-frame image processor. Its architecture should implement various modifications of 2D filtering as quickly as possible, which requires the minimum possible time for image input and the maximum possible parallelization of data processing. The prototype of such an architecture is based on the simultaneous reading of all sensor pixels and the subsequent implementation of the $MLPF$, shown in Figure 2. In this case, the minimum time complexity is ensured, i.e., the minimum time for intraframe processing. This allows us to consider the architecture under consideration as the basis for the future intraframe processors design.
- Secondary (interframe) image processor whose RISC architecture is designed for fast implementation of iterative algorithms. Possible analogs can be the architectures of the well-known family of STM controllers [10].
The image processing video system architecture as part of a mobile real-time system is shown in Figure 3.

Figure 3. Real-time video system architecture.

The main operating mode of the video system is continuous reception of pixel streams from the sensor output, streaming image processing on an array of processing elements (PE), and iterative processing of the results on the RISC controller. The design of the INTRAFRAME PROCESSOR requires the solution of significantly more complex scientific and technical problems in comparison with the INTERFRAME PROCESSOR. Therefore, further attention will be paid to the architectures and physical implementation of INTRAFRAME PROCESSOR.

4. Homogeneous computing environments and functional-oriented processors

One of the effective approaches to the development of INTRAFRAME PROCESSOR is the use of homogeneous computing environments (HCE). Implementation of massive parallelism of data processing in HCE allows to provide maximum performance in intraframe image processing. In addition, homogeneous structures are optimal from the point of view of using the features of semiconductor microelectronics [11], [12].

The HCE is a two-dimensional lattice, at the nodes of which PE are located, each of which is connected to the others only by local connections, forming a lattice graph [11]. All PEs are identical, each of them is a finite state machine that implements a minimal set of operations combined with data transfer in different directions. The PE may include local RAM. The structures of a 2D array and a single PE are shown in Figure 4. The local connectivity of the HCE allows the number of PEs to be increased by a simple docking method, which makes it easier to increase the computing performance.

Figure 4. The 2D-array of HCE and structure of single PE.

HCE is the basis for the development of special-purpose processors for massively parallel data processing. The architecture of such processors and their set of instructions are oriented on the rapid performing of basic digital image processing procedures; therefore, they belong to the class of functional-oriented processors (FOP). Let's call such processors as 2D-HCE-FOP.

There are two types of 2D-HCE-FOP architectures:

- Architectures with one command stream and multiple data streams - SIMD [13]. This type of 2D-HCE-FOP contains the PE array, which executes one general program in clock cycles and
implements a specific algorithm under the control of one instruction stream. The configuration of the PE array does not depend on the implemented algorithm, it is determined by the structure of the data array. For example, for video systems, it is associated with the image geometric characteristics. For this type of 2D-HCE-FOP, data processing optimization is possible only at the algorithmic and software levels, since the architecture is predetermined.

- Multiple Command Stream and Multiple Data Stream architectures - MIMD [13]. These 2D-HCE-FOP contain many PEs that are connected locally and pre-configured to implement a specific algorithm. Each PE executes only one command during the entire time interval for solution of the problem. The PE array configuration is completely determined by the implemented algorithm. The 2D-HCE-FOP architecture of the MIMD type is the result of mapping the algorithm graph onto the HCE cellular space.

5. Two-dimensional-HCE-FOP for intra-frame image processing

5.1. SIMD-2D-HCE-FOP

The development of a processor of this type for the first time in Russia was carried out in the late 1990s by a team of developers, while some well-known projects were prototypes, for example, a GAPP processor developed by NCR, USA [14].

The FOP architecture allows to input pixels at the rate of arrival and process them in such a way as to complete the processing by the time the next frame arrives. The FOP architecture allows reading pixels from the sensor output in real time and processing them in such a way as to complete the processing by the time the next frame arrives. This FOP has two separate interfaces - one is designed to work with a sensor, the other - with a central computer system.

To receive video frames, the FOP has a systolic buffer, which receives the next frame line and provides a parallel recording of all its pixels in the PE array.

Researches of this systolic 2D-HCE-FOP have shown that, compared to other architectures, it provides the maximum rate of intraframe image processing. For example, contouring an arbitrary number of extended images that fall into the plane of the sensor takes only about ten clock cycles. The architecture of SIMD-2D-HCE-FOP is shown in Figure 5.

![Figure 5. SIMD-2D-HCE-FOP for real-time intraframe image processing.](image)

At the stage of developing the video system, the design of the VLSI of the systolic operating unit (SOU) was carried out. The SOU architecture is an 8×8 PE matrix. Each PE contains a bit ALU, a system of switches and RAM with a capacity of 128 bits. In addition, a micro-board with size (60×48)
mm² was developed, on which 8 VLSI chips are mounted, resulting in 512 PE. A full-scale 2D-HCE-FOP consisted of 32 micro-boards. It implemented intraframe image processing at a frame rate (f = 100Hz) and had a mass of 200 g and a volume of 0.3 dm³ at a power consumption of no more than 10W.

5.2. MIMD-2D-HCE-FOP

Architecture of MIMD-HCE-FOP differs from SIMD architectures in the mode of operation. At the preliminary stage, each PE is configured to execute a specific command, and after all PEs are configured, data streams are supplied to the input interfaces and begin to be processed and transmitted from one PE to another throughout the OVS. Data processing in this case is combined with the process of moving them through the PE array. Obviously, in this case, the fastest processing of information is realized, since it is carried out at the rate of input data reception.

In the mid-1990s, the VLSI "MiniTera" was developed, which is a MIMD-HCE. Each PE contains an ALU, a stream multiplier, a configuration register, and a 64-bit stream register. The instruction set of each PE consists of 49 instructions. The structure of PE HCE "MiniTera" is shown in Figure 6.

![Figure 6](image1)

**Figure 6.** The structure of PE HCE "MiniTera".

Each VLSI contains 25 PE. Information channels provide simultaneous data exchange between PE(i,j) and PE (i-1,j), PE (i+1,j), PE (i,j-1), PE (i,j+1) respectively. The topology of the tuned PEs unambiguously corresponds to the structure of the algorithmic graph. Figure 7 shows the structure of the HCE "MiniTera". As a result of the R&D, a model of the MIMD-2D-HCE-FOP was developed on the basis of the VLSI "MiniTera" and a computer workstation was created, on which a number of applied programs for digital processing of signals and images were developed and tested. The workstation provided HDTV procedures with a frame rate of approximately 100 Hz.

![Figure 7](image2)

**Figure 7.** The structure of the HCE "MiniTera".

6. Conclusion

Intraframe image processing is one of the most computationally complex for mobile real-time video systems. Intraframe processing algorithms have the highest possible level of parallelism, which allows achieving high performance of video processors. One of the possible solutions to the problem of achieving the minimum time for intraframe image processing is the use of functional-oriented processors based on homogeneous computing environments - HCE-FOP. Processors with a homogeneous structure allow minimizing the time of intraframe image processing by organizing the processing of pixel streams on two-dimensional arrays of PE. The microelectronic implementation of HCE-FOP confirms the effectiveness of their use as the main blocks of video processors in mobile systems.
References

[1] Jensen J R 2015 Introductory Digital Image Processing: A Remote Sensing Perspective (Pearson Education, Inc)

[2] Lin A Hardware Implementation of a Real-Time Image Data Compression for Satellite Remote Sensing Remote Sensing – Advanced Techniques and Platforms ed B E Ramirez (Rijeka, Croatia: InTech) chapter 18 pp 415–428

[3] Popescu D, Stoican F, Stamatescu G, Ichim L and Dragana C 2020 Sensors 20 817

[4] Zheng J, Lu C, Guo J Chen D and Guo D 2019 ACM Trans. Des. Autm. Electron. Syst. 24 2 Article15

[5] Voevodin V V 1986 Mathematical models and methods in parallel processes (Moscow: Science)

[6] Tsilker B Ya, Orlov S A 2006 Organization of computers and systems: Textbook for universities (Saint-Petersburg: Peter)

[7] Vatutin E I, Zotov I V, Titov V S 2010 Combinatorial logic problems of synthesis of partitions of parallel algorithms of logical control in the design of logical multicontrollers (Kursk: Kursk State University)

[8] Lookin N A 2008 Proc. Fifth Int. Youth Sci. Tech. Conf. (Taganrog: Publishing house TTI SFU) pp 115 – 166

[9] Hopcroft J E, Motwani R, Ullman J D 2007 Introduction to Automata Theory, Languages, and Computation (Pearson Education)

[10] https://www.st.com/en/automotive-microcontrollers.html

[11] Prangishvili I V, Abramova N A, Babicheva E V, Ignatushchenko V V 1967 Microelectronics and homogeneous structures for the construction of logic and computing devices (Moscow, Science)

[12] Bogachev M P 1981 Computer system architecture with a homogeneous structure Homogeneous computing environment. Architecture and implementations ed. V V Gritsyk, Yu N Sulitsky and oth (Lviv: FMI)

[13] Flynn M J 1972 IEEE Trans on Comp pp 948—960

[14] Preston K 1990 IEEE Trans on Aero Electro Sys 26 3 pp 548 – 558