НАУЧНОЕ ПРИБОРОСТРОЕНИЕ, 1995, том 5, № 3-4, с. 13-16

МОДУЛЬНЫЕ МУЛЬТИПРОЦЕССОРНЫЕ СИСТЕМЫ И СЕТИ

УДК 681.325.3:003.23

PERFORMANCE EVALUATION OF THE SM-IMP ARCHITECTURE: A PARALLEL, ETHEROGENEOUS, IMAGE PROCESSING ORIENTED ARCHITECTURE*

© 1995, M. Migliardi, M. Maresca

DIST - University of Genoa, Italy

Image processing is one of the great challenges of computing, and image processing problems have to be solved in different application fields like telecommunications, biomedical instrumentation development, robotics and many others.

In each of these fields, the image processing part is one of the most computationally heavy and time consuming parts. Very often, image processing applications need both to process enormous quantity of data and to perform complex computations. For these reasons a large number of parallel computers dedicated to image processing tasks have been developed. These machines usually consist of a set of processors homogeneous in terms of computing power and can be modeled using a single computing paradigm.

Nevertheless, it is well known that image processing problems are usually structured so that no homogeneous architecture is able to perform well in every task forming a complex application, but some tasks are well suited for a computational paradigm while others are well suited for a different one.

In this paper we describe the SM-IMP system, an etherogeneous, image processing oriented parallel architecture built using commercially available and highly standard computing and interconnection systems and we evaluate its performance using the standard MPEG coding algorithm as a benchmark.

Introduction

Image processing is one of the great challenges of computing, and image processing problems have to be solved in different applicative fields like telecommunications, entertainment, biomedical instrumentation development, artificial intelligence, robotics and many others.

In each of these fields, the image processing part is one of the most computationally heavy and time consuming parts. Very often, image processing applications need both to process enormous quantity of data and to perform complex computations with very strict time limits.

For these reasons, in past years a large number of parallel computers dedicated to image processing tasks have been developed [1]. These machines usually consists in a set of processors homogeneous in terms of computing power, and can be divided in two broad categories: Single Instruction Multiple Data (SIMD) systems and Multiple Instruction Multiple Data (MIMD) systems [2].

Nevertheless, it is well known that image processing problems are usually structured in such a hierarchical way so that no homogeneous architecture is able to perform well in every subtask forming a complex application; on the contrary some subtasks are well suited for the SIMD computational paradigm while others are well suited for the MIMD computational paradigm.

For this reasons we have decided to study a system incorporating both an SIMD subsystem and an MIMD subsystem.

In past years some such systems have already been built [3]. Although they met the requirement of incorporating in a single computing system both an SIMD part and an MIMD part they were based on strictly proprietary hardware and software solutions and so they lacked any degree of openness.

On the contrary our goal was to study a solution that would have allowed us to develop a system both heterogeneous and open.

For this reason we have decided to base the realization of the sm-imp system on commercially available and highly standard building blocks such that future developments would nicely fit into the system and not clash with its structure.

In the next section we will describe the subsystems of the whole architecture, in section three we will describe the standard MPEG [4] coding algorithm, in section four we will analyze the performance obtained with the sm-imp etherogeneous parallel system and finally in section five we will provide some concluding remarks.

***The work described in this paper has been supported in part by reseaech grants from the Europran Economic Community (SPRIT project 8849, SM-IMP), MURST (Ministero della Universita e della Ricerca Scienfica e Tecnologica) and CNR (Cjnsiglio Nazionale delle Ricerche).**

The subsystems of the sm-imp etherogeneous architecture

The sm-imp architecture, can be divided in four subsystems: the SIMD subsystem, the MIMD susbsystem, the SISD subsystems and the interconnection subsystem.

The SIMD subsystem is composed by a MasPar MP1 array processor with 1024 4bit processors. Each processor has 64Kbytes of local memory and the whole array processor share a 12.5 MHz clock signal. The array processor is connected to the whole sm-imp system through a front end DECstation 5000-240.

The SIMD subsystem is very well suited to memory intensive computing tasks. In fact, as each processing element (PE) has its own connection with a local data memory the aggregated memory bandwidth is very high.

On the contrary, as each PE has a four bit architecture, and the clock signal has a low frequency, the aggregated computing power of the array processor is not very high. Thus the SIMD subsystem is not well suited to very CPU intensive tasks.

The MIMD subsystem is a Parsytec Ltd. TIP system. Our subsystem is a PowerPC based MIMD multiprocessor with two computing nodes inteconnected by a 100 Mbytes synchronous bus.

The multiprocessor is connected to the whole sm-imp system through a front-end Sparcstation 10-512.

The MIMD subsystem is well suited to very CPU intensive subtasks as a matter of fact each node has a very high CPU power. On the contrary the low number of nodes gives to the MIMD subsystem a moderate memory bandwidth.

The SISD subsystem is composed of three mid to high end workstations. Two of these workstations are also the front-end part of the SIMD and MIMD subsystems and the third workstation is a Sparcstation 5-70.

This workstation are used by the sm-imp system as file-system servers and to fill in the computing gaps that can be caused by an unbalanced partitioning of a complex application.

The interconnection subsystems is composed by a standard Fiber Distributed Digital Interface (FDDI) ring. The FDDI ring is a fiber optical based communication network that gives a 100 Mbit/s shared bandwidth to the connected nodes.

The standard MPEG coding algorithm

The MPEG standard defines an algorithm for the compression of moving video sequences. The MPEG standard has been developed to provide a common format:
— for the storage of compressed video data on digital memory devices, i.e. CD-ROM, Winchester and optical disks;
— for the transmission of compressed codes on band limited channels.

The so-called MPEG 1 Standard, or Constrained Parameters is targeted to approximately 1.5 Mbps channels.

The code representation, as defined in this standard, allows an high compression ratio, preserving on the other hand a good quality of the whole video sequence.

This compression algorithm, however, is not "*lossless*"; it means that the pixel values are not exactly preserved through the coding-decoding process. Hence, each frame is intentionally degraded (keeping the subjective visual quality acceptable) in order to achieve higher compression ratios.

The compression scheme described in the MPEG standard is based on many techniques adopted in the different phases of the coding process.

In this section we briefly describe each of these phases.

Like the television systems, this standard allocates different bandwidth to the different kinds of visual information. Human Visual System (HVS) highly appreciates the luminance, and thus luminance is coded with the maximum resolution. Since in the HVS, the spatial sensitivity to the colors is less relevant, chrominance values are two times undersampled for each dimension, allowing information reduction without appreciably degrading images.

The term *quantization*, in this context, denotes the process of reducing the already discrete resolution of each pixel value. The difference between actual and quantized values is called *quantization noise*. The HVS is less sensitive to the quantization noise than to other kinds of noise, for this reason quantization performs a good compression and produces a little loss in the quality of frame.

The motion compensation is a technique that exploits the correlation between two different frames of a moving sequence. If these frames are temporally closed each other they are likely to be highly correlated. The idea of the motion compensation algorithm is to represent the current frame by considering the object motion with respect to another frame called *reference frame.*

To compute this motion, according to the MPEG standard, this algorithm analyzes the frames previously fragmented into rectangular subsets of pixels called macroblocks. This analysis verifies if and where each macroblock of the current frame is also present in the reference frame.

The results it yields are the vectors which represent the macroblock motions frame by frame.

The reference frame can either temporally follow or precede the current frame: in the first case a *backward prediction* takes place, a *forward prediction* in the other case. The *bi-directional prediction* is the combination of both previous predictions.

Corresponding to these three kind of predictions three types of frames are defined: if a frame is not reconstructed through any kind of prediction it belongs to the I-FRAME (*Intraframe*) class; if it's predicted through the forward prediction it belongs to the P-FRAME (*predicted*) class: if through bi-directional prediction, it belongs

to the B-FRAME (*bidirectionally predicted*) class.

The Discrete Cosine Transform (DCT) is performed to convert the values of a 8x8 pixel block into the corresponding 8x8 array of spatial frequency coefficients. Generally, most of the energy is clustered in the upper left corner of the array, i.e. in the low spatial frequency components. The DCT coefficient at the (0,0) position is called DC coefficient. It is proportional to the average of the values of all the pixel within the non-transformed block. Since DC coefficients of consecutive blocks are likely to be similar, they are coded by a differential coding.

The other 63 DCT coefficients are called AC coefficients because they refer to non-zero spatial frequencies. These coefficients typically decrease against the increasing spatial frequencies. Besides, a rawer quantization of high frequency coefficients often forces them to zero.

So, if these coefficients are read according to a "*zig-zag scan*", they are likely to originate a stream with long zero sequences, well suited to a run-length coding.

Values obtained from the run-length coding are again coded through an entropy coding based on a sub-optimal Huffman-like code (i.e. shorter codes for more frequent symbols).

Performances of the sm-imp etherogeneous parallel system in MPEG coding

Implementing an algorithm on an etherogeneous parallel system is quite different that implementing it on an homogeneous parallel system. As an example, if the target architecture is an homogeneous SIMD system, the parallelization of the algorithm requires a data parallel approach for the whole algorithm. This process can lead to a somewhat abstract implementation that is unable to use the natural parallelism of the algorithm in all of its parts but forces it to evolve in an artificial way. The same effect can be observed in realization targeted to homogeneous MIMD system that force a control parallel approach.

On the contrary, an etherogeneous architecture allow the application developer to exploit all the parallelism present in an algorithm in its natural form.

In the remainder of this section we describe how we mapped the MPEG coding algorithm on the subsystems of the sm-imp architecture and we analyze its performance.

The most computationally intensive tasks of the MPEG coding algorithm are:

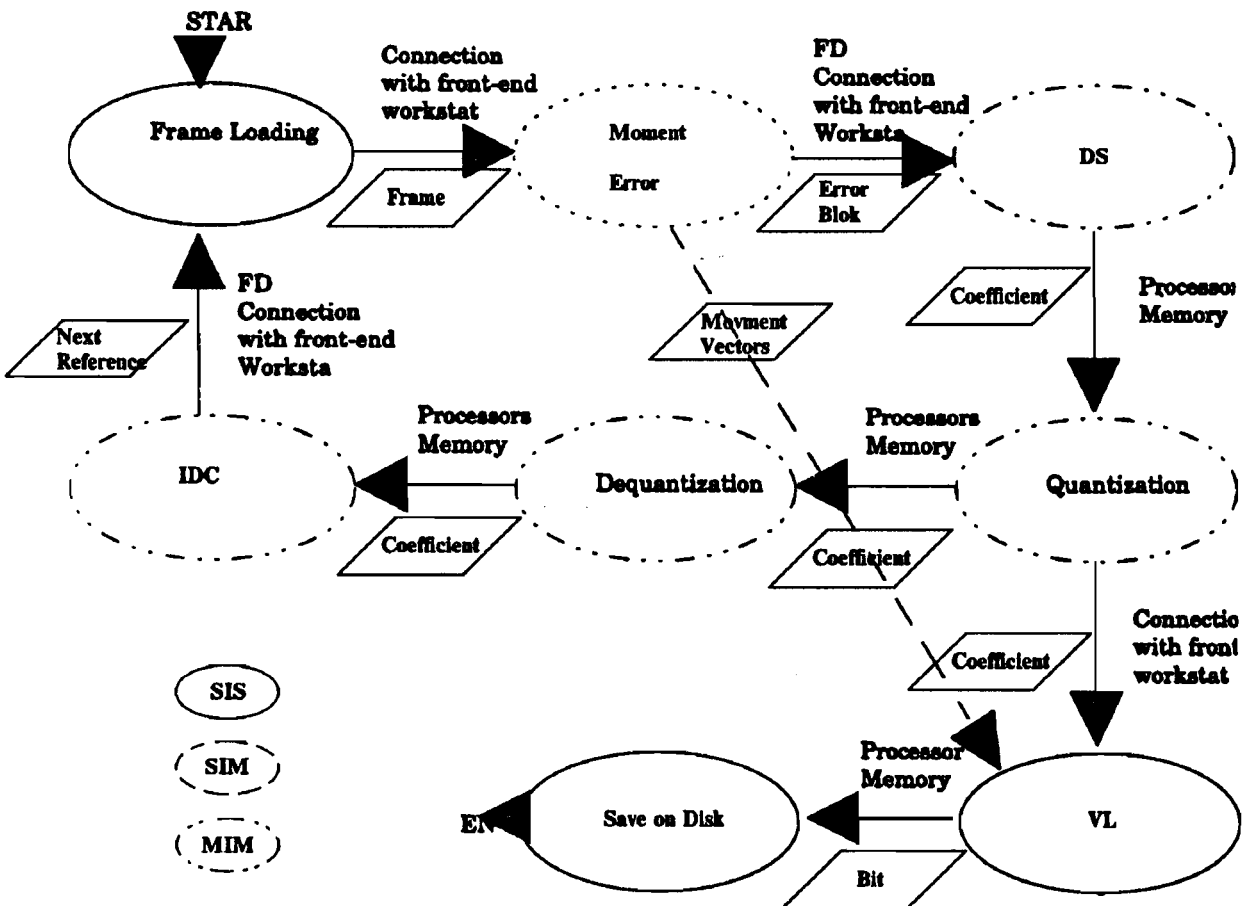

Figure 1 Subtasks mapping for MPEG coding algorithm.

— the motion estimation;
— the DCT and IDCT.

In figure 1 we show the mapping we adopted for this algorithm.

This mapping is driven by the characteristiques of the subtasks in terms of parallelism and in terms of computational requirements.

The motion estimation is composed mainly by small integer operations and memory operations. Besides it lends itself very naturally to a data-parallel implementation. For this reason we have decided to map it on the SIMD subsystem.

The DCT and IDCT are composed mainly by multiplications. Besides they lend themselves very well to a pipelined implementation. For these reasons we have decided to map it on the MIMD subsystem.

Using this mapping, the main bottleneck of the sm-imp architecture in the MPEG coding algorithm is the dimension of the SIMD array processor. With only 1024 4bit PEs the sm-imp architecture is able to code a single macroblock in about 8 milliseconds, a performance that does not go beyond high end workstations. Nevertheless the algorithm implementing the motion estimation is quite scalable and the whole sm-imp scales very well too. Besides the mapping adopted keeps the data streams traveling across the FDDI network quite low and so a more powerful SIMD system will give an immediate boost to the sm-imp system performance.

Concluding remarks

In this paper we have presented the sm-imp system, an etherogeneous, parallel, image-processing oriented architecture built using commercially available highly standard subsystems.

The main features of the sm-imp architecture are:

1. etherogeneity;
2. openness;
3. scalability.

Using the MPEG coding algorithm as a benchmark we have highlighted the fact that the main bottleneck of the sm-imp architecture is the low number of PEs of the SIMD array processor. Nevertheless, the openness and the high scalability of the system would allow to upgrade the SIMD array processor quite easily resulting in a power-up of the whole architecture.

References

[1] J. Nickolls, *The design of the MasPar MP1: a cost effective massively parallel computer*, Proceedings of Compcon Spring 1990, San Francisco 26/2 2/3 1990.

[2] M. Flynn, *Very High Speed Computing Systems*, Proceedings of the IEEE 12, 1966.

[3] M. J. Colaitis, *P3I a multiparadigm video machine*, Internal Esprit Report, LER1 January the 26th 1994.

[4] MPEG-1 Standard (ISO/IEC International Standard 11172-2.

ОЦЕНКА ПОКАЗАТЕЛЕЙ РАБОТЫ SM-IMP АРХИТЕКТУРЫ: ПАРАЛЛЕЛЬНОЙ, ЭТЕРОГЕННОЙ АРХИТЕКТУРЫ, ОРИЕНТИРОВАННОЙ НА ОБРАБОТКУ ИЗОБРАЖЕНИЙ

М. Миглиарди, М. Мареша

Обработка изображений – одна из основных проблем обработки данных. Задачи обработки изображений необходимо решать в различных областях применения, таких как телекоммуникации, разработка биомедицинской аппаратуры, робототехника и многие другие.

В каждой из этих областей – та часть аппаратуры, которая связана с обработкой изображений, является наиболее сложной в вычислительном отношении и требует наибольших затрат времени. Очень часто в задачах обработки изображений требуется обрабатывать огромное количество информации и выполнять сложные вычисления. По этим причинам было разработано большое количество параллельных вычислительных устройств, специально предназначенных для обработки изображений. Обычно эти устройства включают в себя ряд однородных, с точки зрения вычислительных возможностей, процессов, которые можно смоделировать с помощью единой вычислительной парадигмы.

Тем не менее, хорошо известно, что проблемы обработки изображений обычно так структурированы, что ни одна однородная архитектура не способна функционировать одинаково хорошо во всех заданиях, составляющих сложную задачу, но некоторые задания хорошо подходят для одной вычислительной парадигмы, а другие для другой.

В настоящей статье описывается система SM-IMP – этерогенная, ориентированная на обработку изображений параллельная архитектура, построенная с использованием серийно выпускаемых и высоко стандартных вычислительных и соединительных систем, оцениваются характеристики ее работы с помощью стандартного алгоритма кодирования в качестве критерия.