THE PRESENT AND FUTURE OF SIGNAL PROCESSING

VITO CAPPELLINI, PIER LUIGI EMILIANI

The paper reviews many important activites in digital signal processing in particular some topics interesting as a starting point in present and future depth analyses and discussions, with emphasis on implementation problems.

The necessary changes and trends in design and implementation techniques of digital signal processing systems are described, which result from development of related technology with respect to costs and performances suitable for many application fields (image processing, speech recognition, pattern recognition, etc.). A significant new feature of signal processing follows from the fact that the number of arithmetical operations is not quite a correct figure of merit for the evaluation of relative efficiency of algorithms. Instead, other parameters, as segmentability of the algorithms and, particularly, the amount of communication between building blocks is becoming of increasing importance. From this point of view the following, three different classes of architectures are discussed in the paper:

1. Unconstrained architectures. In this class, at least in principle, no constraints of control or arithmetic are imposed by the available components. This means that any architecture can be designed to be adapted to the chosen algorithm.

2. Constrained architectures. This class consists of the implementations based on (i) general purpose computers, minicomputers and microcomputers, (ii) digital signal processors.

3. Highly parallel architectures. In this class several approaches to the implementation of parallel processors with different complexity and performance levels can be considered.

1. INTRODUCTION

Digital signal processing (1-D and 2-D) is assuming a new role in research and applications due to several circumstances:

- the present impact of 1-D techniques in consumer electronics (mainly in digital audio) and in telecommunications, where, due to the high volumes of production, complex algorithms can be efficiently implemented in VLSI, and the foreseen extension of 2-D techniques into the video area (digital video disks, interactive video, etc.);

- the increasing importance of images in many applications (e.g. communications, radar-sonar, remote sensing, biomedicine, office automation, moving object recogni-

tion and robotics) with the corresponding emphasis on the development of efficient algorithms, of workstations for simulation and of parallel organizations of processors for real-time processing of images and sequences of images;

- the development of narrow band telecommunication applications, where data, voice and images can be transmitted in digital form (ISDN), which creates a potentially huge market for all the techniques and technologies for A/D and D/A conversions and signal processing;

- the developments in the direction of broad band communication networks (IBC), with the corresponding increase of importance of picture and graphics (TV, HDTV).

This is made possible by the parallel development of related technology (circuitry and computer aided design facilities), with costs and performances suitable for many application fields. The following examples can be considered:

- the impact of DSP chips, which, due to their decreasing costs and increasing performances, contribute to the application of digital techniques also where volumes of production are not compatible with custom implementations (e.g. in measurement instrumentation, biomedicine and industrial nondestructive tests);

- the development of new technologies (standard cells, gate arrays) for the production of semi-custom circuitry, in applications where full custom VLSI are not economically viable;

- the increase of performances of available technologies (e.g. CMOS), the development of new high speed technologies (e.g. GaAs) and the development of parallel processing approaches.

Obviously it is not possible to review in the short space at our disposal all the possible activities in signal processing. We will try only to point out some topics, which can be interesting as a starting point for future in depth analyses and discussions, with main emphasis on implementation problems.

2. REVIEW OF SIGNAL PROCESSING TECHNIQUES

Many techniques are now available to process signals and images.

Several forms of transforms have been defined (Fourier transform, Hadamard transform, Walsh transform, Haar transform, slant transform, Karhunen-Loève transform, SVD – singular value decomposition), which are used in measurement applications, in data compression and as building blocks of complex processing operations. Fast algorithms have been defined, excluding only the Karhunen-Loève transform. Of particular interest in this field (mainly in the 2-D case) is the investigation of implementation architectures, which allow an efficient use in real-time applications.

Digital filtering is a second processing technique, of fundamental importance in 1-D and 2-D applications. The basic theory of design and implementation of corresponding algorithms is now available [1, 2] and some applications are now reaching the level of consumer electronics. Open problems exist in the design of 2-D infinite impulse response filters (IIR), even if from the application point of view the main emphasis in 2-D is on finite impulse response filters (FIR), which can be designed to have linear phase (which is required to have no phase distortion). An example of 2-D FIR is shown in Fig. 1.

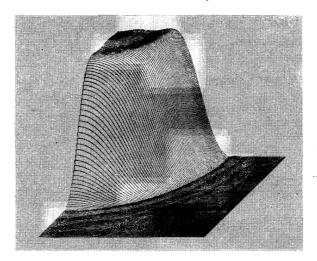


Fig. 1. Example of spatial frequency response (amplitude in one quadrant) of a circular 2-D FIR low-pass digital filter (16×16 coefficients).

Most of the interest is now in the definition of efficient implementation structures, while new components are appearing in the market for their realization, as it will be shown in the following.

However, 2-D general filtering which to be efficiently implemented requires operations in the frequency domain, that is global operations on the image, is often too demanding from the point of view of computation and some interest is in local space operators, that is in low complexity 2-D systems, which process small blocks of data (image samples) in the space domain. Smoothing operations can be simply obtained computing the mean value on the block data, while enhancement is possible by evaluating, for instance, the difference of data along lines and columns. Another very important class is represented by edge detectors, which extract boundaries and edges in the processed images. Local space operators are important mainly because of their simplicity of use, ease of implementation and possibility of use in real time. This is due to the small number of operations involved and to the local access to memory [3].

Of increasing interest are also 2-D local space operators performing nonlinear filtering operations. This is a common trend in many applications where nonlinear filtering (e.g. median filtering) is becoming very important.

Data compression (that is the reduction of redundancy in signals and images) is now particularly interesting for several applications, as the diffusion of narrowband communication networks, with the inherent necessity of efficiently coding speech, audio and video information, and the need in many application fields (e.g. biomedicine) of setting up large data banks, where images are stored in digital form. Several data compression schemes have been proposed for 1-D and 2-D signals, based on adaptive sampling, prediction, interpolation, filtering, encoding and the use of transforms (Fourier, Hadamard, Walsh, Haar, Karhunen-Loève). The new large scale applications have as a consequence the emphasis on efficient algorithms and implementations [4].

Digital filtering (mainly decimation and interpolation) is the basis of algorithms for geometrical transformations, which are very important in 2-D processing for presentation problems and in applications where images from different sources have to be correlated (e.g. in remote sensing, where often images with different resolution have to be compared).

Speech recognition and several image processing applications (e.g. pattern recognition, artificial vision and so on), where signal processing contributes to the fulfillment of complex goals, are examples of the new multidisciplinarity, which is required in future applications. Knowledge based procedures can be used to decide the best strategies for processing. Measurements made using digital processing techniques can be integrated with artificial intelligence concepts to analyze the results.

As an example the application of image restoration techniques can be considered. Many different algorithms have been proposed for different classes of images. But if one of these techniques is applied to a class of images different from the one for which it was designed, the obtained results are often unacceptable. This is the case also when in a single image different segments with different characteristics are present. Therefore a knowledge based procedure could be useful, which, starting from the identification of segments with different characteristics, would decide and apply the most suitable restoration procedure to the different segments, thus trying to optimize the overall result. This implies the formalization of decision criteria for the selection of processing techniques as a function of the characteristics of the signal and/or image to be processed, of the algorithms available, of the cost of processing.

Another very important application field, where the above considerations are valid, is pattern recognition, in which useful patterns and configurations are extracted, recognized and classified. Digital signal processing can play a very important role in the preprocessing of signals and images (in particular to reduce the noise) and in the extraction of the features of the patterns to be recognized, or to be given as inputs to classifiers.

A final important problem in signal processing is currently requiring a multidisciplinary approach, with contributions from network theory and, probably, incorporating again concepts from artificial intelligence. This is the optimization of communication among processors or processing units which constitute the implementation structure.

In the case of locally connected systems it is necessary to incorporate the problem of communication in the study of the building blocks, to be used in the setting up of the entire structure. In the case of networks of processors, it is necessary again to take into account these considerations in the segmentation of the algorithms so as to minimize the communication between the processing nodes and it is necessary to optimize the transfer of information to minimize the time when processors are idle, waiting for data to be processed.

3. IMPLEMENTATION ARCHITECTURES AND COMPONENTS

Several options are now available for the implementation of signal and image processing operations, which cover important application fields with different levels of prices and complexity [1, 5].

It is important to observe that the availability of efficient digital processing algorithms and the increase of performances of digital circuitry increase the number of application fields, introducing the need of a careful selection of the technique to be used also as a function of the implementation structure and of the possibility of their segmentation.

This is a very important new feature of signal processing, which will be predominant in the next future.

Most of the work in the design of signal processing algorithms was based on the following assumptions:

- that the implementation had to be done on computers or computer like architectures;

- that the most demanding part of the effort in the implementation was in the arithmetic computations and therefore their number had to be minimized.

The situation started to change with the diffusion of image processing, because, in conventional systems the time to access the disk was in many cases predominant with respect to the arithmetic processing time. This is now the rule, both with DSPs (digital signal processors) and special purpose architectures, where the number of arithmetic operations is no more a correct figure of merit for the evaluation of relative efficiency of algorithms.

Particularly with parallel implementations, other parameters, such as segmentability of the algorithms (i.e. the possibility of using simple and compact sub-units for processing) and particularly the amount of communication between building blocks is becoming of particular importance. But also using DSPs, algorithms which fit completely in the chip are much more efficiently run than algorithms, which have to access continuously the external memory. In this case the interest is for algorithms that are compact from the point of view of access to memory, i.e. that can be implemented using adjacent subsets of the data to be processed. As an example FFT (base 2, base 4) is not very suitable in its standard form, because the butterfly algorithm has to access all the memory at every iteration.

For clarity it is interesting to use a simple classification of possible architectures, even if some overlap exists between the different classes.

a) Unconstrained architectures

In this class, at least in principle, no constraints (in terms of control or arithmetic) are impossed by the available components, obviously excluding speed limitations. This means that any architecture can be designed to be adapted to the chosen algorithm.

The usable components are:

- standard LSI components;
- multipliers/accumulators;
- bit slices (to be used in a microprogrammed architecture).

The implementable algorithms are:

- standard signal processing algorithms, where the designed architecture can be matched to the chosen processing structure;

- simplified algorithms (as an example all the filtering and DFT algorithms based on simple multipliers - power of two multipliers - can be considered);

- distributed architectures, where a trade-off between operations and other resources (e.g. memory) is looked for (ROM-based second order section can be considered as an example).

The main fields of application of these techniques are:

- medium to high speed applications (e.g. radar);

- low volume production, when the development of dedicated VLSI is not convenient.

Obviously in the case of volume production, this approach can be transferred into the implementation of VLSI chips (full custom, standard cells, gate arrays and so on), where in principle any primitive can be defined to be used in constructing the architecture.

As already discussed above general purpose algorithms are usually optimised from the viewpoint of minimisation of number of operations and/or of memory. In hardware implementations this approach to the design of algorithms is not the most convenient. Even if the control of the hardware can be constructed to match the structure of the algorithm, the implemented control units become, unfortunately often very complex. They tend to become the most demanding part of the system and to enforce the use of wide and long communication buses. This second circumstance can become the limiting factor in operation speed, even if the problem can be solved increasing, when possible, the number of pipeline levels in the structure. Therefore it appears very important in this case to review the algorithms in order to increase the efficiency of the control.

From the point of view of implementation chips, the older circuits were simple and assembled with repetitive cells (multipliers, adders, other logical elements), with poor programmability.

The chips of the second generation integrate much more elements (ALUs, address generators etc.) and have a large possibility of microprogramming.

The new chips have characteristic near DSPs and by integrating all the arithmetic function within a single chip, the speed of the corresponding multiple chip systems can be very high. As an example the Analog Devices ADSP-2100 can be considered. The designers have chosen a completely external program and data memory, to save silicon area on the chip. However the access to the external memory is very efficient and has no speed penalty, with a cycle of 8 MHz. The chip (16 bits data and 40 bit accumulator) contains ALU, multiplier/accumulator, shifter, data generators, program sequencer and cache memory. The process is CMOS $1.5 \,\mu$ m. The times for some typical algorithms are the following: 8 ms for a 64 taps FIR and 7.2 ms for a 1024 points complex FFT.

Floating point elements (adders, multipliers and higher precision chips are also proposed. They find wide applications for radar and military electronics. Weitek WTL floating units, AMD 2933 floating point processors, Analog Devices floating point ALU and multiplier, TRW floating point are examples of these components. They are pipeline floating point operators (ALU and multipliers) and are direct improvements of the standard fixed point operators. Even if they need an accurate overflow control, they allow simpler implementations of very long FIR filters or FFT transformers.

The possible technologies include bipolar, CMOS, up to gallium arsenide components.

b) Constrained architectures

This class consists of the implementations based on:

- general purpose computers, minicomputers and microcomputers;
- digital signal processors.

So far, the resulting systems are normally usable in low frequency applications, even if new families of DSP and the technologies are increasing these limits.

Obviously, any kind of algorithm can be programmed in the chips, even if their structure often appears more suitable for some particular class. When speed is important and the maximum throughput of the processor, in terms of signal processing operations, has to be increased, the revision of the algorithms to simplify their control becomes very important. In signal processing chips the arithmetic operation time (multiplication and accumulation) corresponds to the machine cycle, as any other operation. Therefore the complexity of control and of access to memory are to be very carefully considered, when a processing algorithm has to be chosen.

In addition, the circumstance has to be taken into account that most of the DSP have different performances, if the data to be used are in the internal memory or in an external memory.

The above considerations have some important consequences, as:

- need of algorithms with simple control;

- need of algorithms with simple access to memory;

- need of algorithms which are segmentable, so that simple subtasks can be performed with all the necessary data in memory or by different DSP, which work in parallel;

- need of efficient programming supports, which also can automatically generate straight line implementations of the algorithms, with a trade-off between speed and memory.

The prices of the first DSPs on the market were relatively low, but not sufficiently low for the target devices (like modems, speech synthesis chips etc.). This is the reason why success was relatively low. Their deficiencies for more complex operations are now recognized and the new chips proposed have better characteristics (fast access to the external memory, presence of various input system, also serial, very high speed up to 50 ns, presence of floating point unit, facilities for multi CPU systems), so that they are becoming more interesting for applications. CMOS is the most used technology.

Their main features are:

- programmability and support: programmable by means of standard software tools (assembler, debugger) and high level languages;

- target applications: general purpose civil applications (few chips with military specifications). In multichip architectures a large bandwidth may be possible;

- limitations: not clear cost/performance ratio in very large scale applications:

- success: in growth, now a large amount of new devices is being presented (not yet available on the market).

As an example, the Texas TMS 320/25 can be considered. The process is CMOS (1.8 µm), the cycle is 100 ns. It integrates a MUL/ACC (32 bits), a 544 internal memory for data and program, auxiliary registers and also a serial input/output. It works also with external memory. However to maintain the maximum cycle speed without wait states fast memories (50 ns) are necessary. A 256 complex points FFT requires 3.44 ms in looped code and 1.5 ms in radix 4 and straight line code.

c) Highly parallel architectures

In this class several approaches for the implementation of parallel processors can be considered with different complexity and performance levels, as:

198

- parallel combinations of arithmetic units, concurrent to the implementation of complex algorithms;

- parallel combinations of DSP, which perform different tasks of a complex algorithm or different primitives of a complex signal processing procedure;

- systolic arrays;
- wavefront arrays;
- data flow computers.

In particular the last three classes require an in depth revision of all the known algorithms, which have to be transformed so as to have localized interconnections in space and time. Mapping algorithms have been presented to transform classical signal processing operations (as linear filters) into partitioned algorithms with spatial and temporal locality. Chips for parallel architectures (like systolic and data flow components) are generally very specialized and fully custom designed. The regularity of the approach to the signal processing applications matches very well the needs of VLSI integration. Some effort is being devoted in many laboratories to produce standardized subsystems of the previous type, so that it is possible to implement large scale systems with regular cells.

This is a fairly new field of activity and it is impossible to forecast the commercial success, even if the use of these components appears the only colution for apparatures in high speed applications. As in the previous case, the problem is how fast the design and production technologies grow up, so that this type of component can be cost effective.

As example of the components, which start to appear on the market, the NCR chip (systolic array processing chip, 6X12 processors, cascadable, CMOS 3 μ m, clock 100 ns, 500 mW) and the NEC chip (data flow DSP, 16 bits data, 200 ns, NHMOS) can be considered.

d) Specific application processors

VLSI chips designed for specific applications are diffused in large scale production (telecommunications, echo canceller, modems, speech coders, signal coders, etc.). Speech synthesizers have, for example, large applications on the market.

In general technologies, architectures and algorithms have been optimised to increase the performances in specific applications. The chips are normally not programmable, even if it is possible to change some of the control parameters.

Some of them, as the ones implemented for transversal filtering applications, have large potential applications, even if so far they are not very diffused on the market.

Target applications and success appears similar to that of custom VLSI.

Examples of special chips for signal processing are the TRW TDC 1028 (digital FIR/correlator, 10 MHz clock, 8 cells, 4 bits data/coeff., 12 bits sum, bipolar) the IMS A100 by Inmos (full 16 bits, 32 stages, transversal filter, with coefficients selectable as 4, 8, 12, 16 bits wide and a data throughput up to 10 MHz, with 4 bits coefficients) and the Digital Filter Processor by Zoran.

4. SIGNAL AND IMAGE PROCESSING

General signal processing algorithms (digital filtering, correlation, convolution, Hilbert transform, fast Fourier transform, adaptive filtering, waveform generation...) have now a practical impact in several application fields. However, what is really new is that some of them are reaching the level of large scale applications, instead of being connected to advanced cost insensitive activities, to special instrumentation and so on.

Moreover a large emphasis is now laid on image processing. Multimedia presentation and interaction methodologies, future foreseen implementation of broad band communication networks and recent industrial developments (e.g. robotics) suggest that this interest will increase in the future, with possible extension to multidimensional signals (e.g. sequences of images).

In audio systems (digital audio) digital signal processing has reached the market of consumer electronics, with the wide diffusion of new digital media to distribute music (compact disks, digital audio tapes) and with the integration of audio with images (digital video disks). Two main directions of activity are foreseen. The first is on the user side. New audio processing equipments will probably be made available in the near future, to set up completely digital home processing systems, with D/Aconversion only at the input of power amplifiers. The second is in the development of completely digital production chains, which start from the recording of music and include all the steps up to the final production.

An integration with image processing techniques is also foreseen, to produce audio and video home processing systems, including image storage and processing operations, as image freezing, zooming and simple enhancement techniques.

Speech processing (coding, synthesis and recognition) is a parallel very important application field. Speech compression is necessary for transmission through narrow band (low bit rate) communication channels and for speech. Speech recognition is particularly important, due to its potentiality as a natural man-machine interface. Unfortunately, so far results are not completely satisfactory and new approaches are under study, which try to incorporate artificial intelligence concepts and techniques.

Telecommunications benefit from these new methods for dealing with audio and video information. Adaptive equalizers, codecs, high speed modems, digital filtering to define bandwidth and reduce noise, spread spectrum techniques are now normally dealt with using digital approaches. The diffusion of narrow-band digital channels (ISDN) increases the emphasis on efficient data compression techniques in the audio, speech and video transmission. Efficient image compression techniques are in fact very important in several applications, when real time transmission is not mandatory and ISDN can be used to distribute visual information. The emphasis on data compression (where signal processing can play a very important role) is caused also by the need of efficient archiving systems (e.g. in the biomedical and art areas), where very large archives of images and signals have to be implemented.

Radar and sonar processing is another very important field of application of digital processing. The developments in technology have increased the processing bandwidths, so that they are now compatible with the corresponding applications. Complex algorithms can be implemented in real time to reduce the noise at the receiving unit and to extract useful information (echoes from the targets). In particular in radar systems, digital target indicators (MTI filters), digital delay line cancellers (to eliminate fixed target components in Doppler information extraction) and digital matched filters are now employed.

In the biomedical area EEG and ECG signals have already been processed for several years (due to their narrow bandwidth), using digital filtering, averaging, fast Fourier transform and so on, to analyse their frequency content and/or to reduce noise. More recently image processing techniques became important due to the diffusion of many new diagnostic methodologies based on biomedical images [6]. Ultrasounds, infrared images (Fig. 2), nuclear magnetic resonance (NMR), computer aided tomography (CAT) (Fig. 3), digital radiology, nuclear medicine are new diagnostic methods, which base most of their real impact on complex numerical algorithms for extracting information from raw measurement data and use sophisticated presentation formats.

Image processing and, in general, two-dimensional and multidimensional signal processing is becoming important also in many other applications, as remote sensing an industry. In remote sensing [7] many images and maps are currently collected by platforms aboard aircrafts and satellites through different sensors (optical cameras, multispectral scanners, microwave radiometers, side-looking radars). These images and maps have, in general, to be processed to improve their quality (geometric and sensor corrections, noise reduction, enhancement) and to obtain final useful results (e.g. extraction of specific regions and land-sea areas for agriculture investigations or water resource monitoring). 2-D digital filters can be applied to smooth the image data (by means of low pass filtering), to perform a space frequency correction or to obtain enhancement (by means of high-pass, band-pass and parabolic filtering), extracting also edges and boundaries (Fig. 4). Seismic data analysis can be considered as a part of remote sensing and has an enormous importance in many problems of extraction of information from measurements made in the earth with acoustic media to identify natural resources.

In the field of moving object recognition and robotics, data measuring the position of the objects or images taken on the object can be filtered by means of 1-D and 2-D digital filters, to reduce acquisition noise and enhance the objects in the background. High-pass or derivative filtering can further extract the edges defining the objects of interest (Fig. 5) [8].

Finally digital signal processing became very important in instrumentation. Spectrum analysis, digital filtering, phase-locked loops, averaging, arbitrary waveform generation, transient analysis are particularly important measurement in many measurement applications and digital instrumentation is widely becoming available.

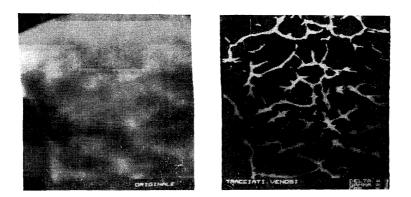


Fig. 2. Example of digital processing of a biomedical IR image: a) original; b) result of application of stretching and edge detection.

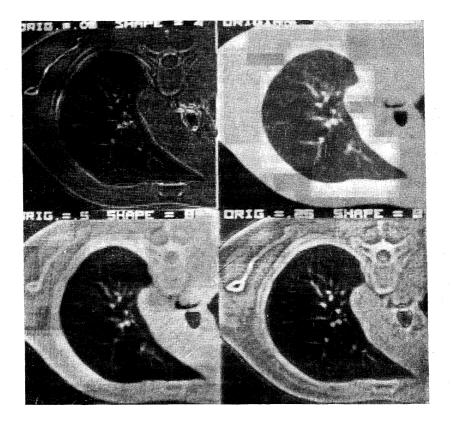


Fig. 3. Example of processing of computer tomography image (top right) by means of a 2-D digital filter of parabolic type: three types of parabolic filters are used producing different enhancement results.

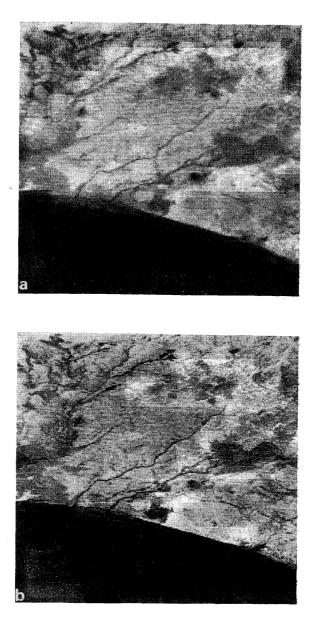


Fig. 4. Example of the application of a 2-D digital filter of high-pass type to a LANDSAT-C image (North Africa): a) original image; b) filtered image.

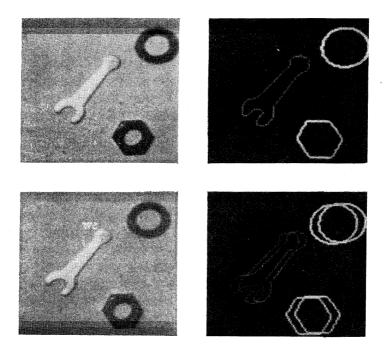


Fig. 5. Example of moving object recognition, using preprocessing digital filtering and boundary FFT matching: two consecutive frames are shown with input digitized images at the left and recognized objects at the right (each object is recognized with different colour, here appearing as a different gray-level).

5. CONCLUSIONS

As appears from the above synthetic considerations, the area of digital signal processing is very important at the present time and yet more in the near future.

Practically, with the fast improvement of digital signal processors (increased processing speed, size reduction and lower cost), near all possible applications can be covered extending from low to high frequency signals and from static to dynamic images. Indeed, a continuous trend is represented by the multidimensional signal processing: i.e. more and more 3-D processing algorithms are currently used (as in communications, remote sensing and robotics).

Looking forward to the long term future, optical, signal processing will become very important in particular for increasing the processing speed, being performed not only in analog form (as available also at the present time), but also in digital form (with the implementation of an optical digital computer).

(Received June 21, 1989.)

204

- [1] V. Cappellini, A. G. Constantinides and P. L. Emiliani: Digital Filters and Their Applications. Academic Press, London 1980.
- [2] V. Cappellini and P. L. Emiliani: Two-dimensional digital systems and applications: The state of Art. In: Multidimensional Systems (S. Tzafestas, ed.), Marcel Dekker Inc., New York 1985, Chap. 1.
- [3] W. K. Pratt: Digital Image Processing. Wiley, New York 1978.
- [4] G. Benelli, V. Cappellini and E. Del Re: Data compression techniques. In: Data Compression and Error Control Techniques with Applications (V. Cappellini, ed.), Academic Press, New York 1985, Chap. 3.
- [5] V. Cappellini and E. Del Re: Hardware implementation of data compression and error control techniques. In: Data Compression and Error Control Techniques with Applications (V. Cappellini, ed.), Academic Press, New York 1985, Chap. 3.
- [6] V. Cappellini and A. N. Venetsanopoulos: Some high-efficiency two-dimensional digital filters with applications to biomedical image processing. In: CAR'85, 1985 (H. V. Lemke, M. L. Rhodes, C. C. Jaffee and R. Felix, eds.), Springer-Verlag, Berlin-Heidelberg-New York 1985, p. 485.
- [7] V. Cappellini: Internat. J. Remote Sensing 1 (1980), 175.
- [8] V. Cappellini and A. Del Bimbo: Digital processing of time varying images. In: Issues in Acoustic Signal/Image Processing and Recognition, S. Miniato, Italy, 1983 (C. H. Chen, ed.), Springer-Verlag, Berlin-Heidelberg-New York 1983, p. 283.

Prof. Vito Cappellini, University of Florence and Research Institute of Electromagnetic Waves, Nat. Res. Coucil, Via Panciatichi 64, 50127 Firenze. Italy.
Dr. Pier Luigi Emiliani, Research Institute of Electromagnetic Waves, Nat. Res. Coucil, Via Panciatichi 64, 50127 Firenze. Italy.