CUBICORT: A hardware Simulation of a multicolumn model for *3D Image Analysis, understanding & compression for Digital TV, HDTV & Multimedia.*

P. LERAY (1) . Y. BURNOD (2)

(1) Head, Advanced Research in Image Processing Laboratory, CCETT: Neural Networks Project
Centre Commun d'Etude de Télédiffusion et Télécommunications - BP 59, 35510 Cesson-Sévigné, France.

(2) Head of Computational Neurosciences Group, Univ of PARIS VI

Abstract:

We describe here simulation elements and results of a new kind of 3D Vision Machine, for pre-processing in 3D Object & movement analysis, using the biological concept of the Cortical Column Paradigm in the primary visual area. The target machine is primarily dedicated to image compression for Telecommunication of TV, HDTV, and 3D TV, but can also be used for automatic modelling, digitizing, robotics or medical applications.

Introduction:

3D Image Analysis has been studied in some depth over the years, using algorithmic procedures. Some interesting results are now being achieved, but the problem as a whole (I.E with natural, unknown 3D scenes) remains unsolved. However the Human Visual system is able to analyse and extract 3D objects and movements. Hardware implementations have been proposed for artificial retinas. We present here hardware simulation elements and results to map the properties of the primary visual area, which could be used in a general way for pre-processing, robotics, image analysis, image synthesis (i.e. for digitizing images), and for Telecommunication & storage / retrieval problems such as image compression. As NeuroScience knowledge is constantly improving, it now seems to be possible to design a VLSI chip set which could perform in real time some of the main functions of the primary visual cortex.

Such a machine is currently simulated at the CCETT and is based on a model of cortical columns [Burnod 88] and a 3D real time image synthesis machine [Leray 81; Leray 85].

This machine consists of 4 layers defined by their Primary Index:

1: Pixels, 2: Zones (or Hyperpolygons: which consists of a set of perceptually homogeneous zones of pixels.), 3: Objects and 3: Prototypes.

1/ The Pixel Level:

is a pre-processing, multi-scale pyramidal hardware which computes spatio-temporal filtering models of primary visual areas (V1 in particular) . In this paper, an extended concept of the conventional pixel is called Hyperpixel: In the brain, images received on the retina are transformed with elementary functions or filters in the cortical maps, which generate other information. Here the information extracted from a given retinal zone (receptive fields) is an hyperpixel; A set of Hyperpixels is an HyperMap;

These elementary functions which are biologically performed in the visual cortex can be summarized as: i Spatial Functions: 2D Gabor Filter , (First and second order derivatives) and ii temporal functions (temporal derivatives or correlations between like features shifted in time and sopace). Each function is computed at 5 spatial resolutions, and 5 temporal resolutions. Spatial & temporal resolutions are directly linked and are simultaneously decreasing.

-Suppose we define a N*N pixel image on the input layer: 2 images at t & t+1 are stored, because high spatial resolution does not need a large number of successive images.

-At scale s, we will have $N/4^s$ hyperpixels stored for each image. and the resolution on the temporal axis will be 4^s * 33 mS.

At this level s, we store 4^s images for temporal information.

Each hyperpixel can be extracted according to different resolutions with the use of a special addressing mode in one memory cycle:

At each level, effective address is shifted by 2 in both X and Y direction. Each processor can access a block of hyperpixels and blurr the less significant bits for lowest resolutions. A global symmetry is achieved between X, Y resolution, and Time. Adjustable masks are stored at each level of the machine, and for each processor. A specific training process can be used in order to adapt each mask interactively according to the lower hyperpixels' activities.

In this respect, masks can be modified and considered as synaptic weights.

At this level, a superviser scans activities, and enables masks at the upper levels to be changed.

The first layer uses Gabor oriented Filters (Product of a Gaussian by a sinus) and Laplacian filters.

The results are stored in a set of Hyperpixels.

From Gaussian and Laplacians filters we extract at this level curvatures, which are directly related to the Attention Points: These points correspond to highly informative zones of the picture.

Let: L = Laplacien; R = $(d^2G/dx^2 - d^2G/dy^2)/2$, Theta1 = 1/2Arc cos $(dG/dx)^2$/Gradient); Theta2 = 1/2Arccos $d^2G/dxdy/R$;

The Curvature C is defined by:

C = (L+R*cos(2*Theta2 - 2*Theta1 + PI/2))/Gradient

Attention points are defined by Hyperpixels having a strong C activity .

At this level perceptive grouping is also performed , which gives a first estimation of homogeneous textures. Perceptive grouping can be modelled by an iterative process taking into account the successive modification of the model's filter columns activity. At each image at time t, the activity Aj of the Hyperpixel with the index j can be expressed as follows:

$$A_j(t) = A_j(t-1) + \Sigma_{i=1,4} (\Phi(C,Ci)^* A_{ji}(t-1))$$

Φ: function of 2 columns

2/ The upper levels, which analyse images with unknown coefficients:

- The hyperpolygon (Zone) level,

The Object level, which is set of hyperpolygons. At this level, objects are viewpoint dependent.

- The prototype (or Label) level which is a collection of objects viewpoint independant.

3/ A reconstruction module which draws 2D or 3D areas.

At each level of the system, a reconstruction process is used; (feedback image synthesis) Reconstructed Image is compared with analysed image.

General Scheme of a Coding system:

At each level the Cubicort machine is composed of:

- An input Hypermap for the input Image (Feedforward Process)
- An Hypercolumn for the Hyperpixel process, and directed by the Primary Index

Basically the Hypercolumn is built with a set of Multiplier/Accumulators, and a learning block which concatenates by a bi-linear function the different results obtained by other Column Index and other layers;

- An output HyperMap for the output (reconstructed) Image (Feedback process)
- The associated Hypercolumn.

Such a system can be used as a coding machine, a similar machine beeing used for the decoding process; the links between them being only the activities variations of the different hyperpixels. Such a system can combine low and high level analysis and reconstruction.

But with this approach, we do not need the complete implementation of the 4 levels: For example, the filters used at the first level are roughly equivalent to the DCT, Sub-band or Wavelet Coding Algorithms.

Moreover, perceptive grouping and Order of Priority of informative zones defined by attention points already constitute a key element for coding enhancements: For these zones we can use standard coding schemes such as D.C.T. (MPEG), according to the speed of the region: the smaller the speed, the smaller the compression ratio.

Hardware Simulation: First results

1 - Pre-processing hardware simulation: The goal is to accelerate the filtering process by the use of hardware Multipliers/Accumulators. We have now entirely simulated the first level, with

results concerning the preceptive grouping and attention points. (Such results will be joined in the final paper).

Conclusion:

These simulation results are a first step towards knowledge based, Multiscale Vision Analysis systems which could be used in ultra-high compression schemes as for future telecommunication standards with MPEG 4. Attention points and perceptive grouping is a new key element for giving relative informative importance to zones inside images. The Cortical approach is more a generalization of DCT and Transform Coding Schemes, with a fine tuning, image dependent, of the filtering coefficients, and a cooperation of networks computing various kinds of visual information.

As a symetric approach is used for Analysis/Synthesis, such a machine could also be used for object digitization in Image Synthesis techniques.

REFERENCES

Alexandre, F., Guyot, F., Haton, J.P., Burnod, Y (1991) The cortical column, a new processing unit for multilayered network. Neural Network, Vol. 4, pp 15-25.

Burnod, Y. (1988) An Adaptive neural network: the cerebral cortex. Book: 400 pps. Masson; Prentice Hall.

I. Otto, E.Guigon, P. Grandguillaume, L. Boutkhil & Y. Burnod. Direct and Indirect Cooperation between Temporal and Parietal Networks for Invariant Visual Recognition (1992). Journal of Cognitive Neuroscience, 4 (1) 35-57

Leray, P. (1981) A 3D Syntheic Imagery Generator in real time:*Image Generation/display Conference II* Af Human Resources Laboratory, Phoenix, AZ pp 78-89.

Leray P. (1985) A 3D animation system Datastructures for Raster Graphics *Data Structures for Raster Graphics* Springer-Verlag pp 165 - 171.

Leray P. (1987) Towards a Z-Buffer and Ray-Tracing Multimode System based on Parallel Architecture and VLSI chips in *Advances in Computer Graphics Hardware I* Springer Verlag pp 141 - 145

Leray P. (1990) Modélisation et architecture de machines de synthèse d'image pour la représentation et le rendu d'objets tri-dimensionnels sur écrans graphiques à balayage. *University of Paris II*, Doctorate 1990.