# New advances in Neuro - Biological Simulation of the Visual Cortex for Real World Computing,3D Image Analysis and 3D object Digitization

Pascal Leray

France TELECOM/CNET/DIH
Centre National d'Etudes des Télécommunications

## Abstract

1 – IMAGE BASED DIGITIZING:

3D Rendering tasks are now made easier thanks to the numerous 3D graphics cards available on the marketplace. But automatic 3D objects digitizing is the key challenge for today 3D computer graphics development. The goal of this paper is to present research results and new avenues towards direct 3D digitizing with multi-views cameras, in order to design 3D objects as simply as 2D scanning.

At present, classical epipolar techniques and disparity extraction don't give quite convincing results. Moreover, 3D laser based digitizing technique provide a huge polygon number on their outputs. Most objects and scenes are now digitized by 3D classic modeling techniques, which is a very old, expensive, not automatic, and time consuming method.
Recent advances in low level vision analysis gave interesting results related to Automatic neurofocalisation, and perceptive grouping, mainly thanks to Professor Burnod researches in the field of Neuro Visual hardware models.
These techniques are based on multi scale hyper complex filters such as Gaussian, Laplacian derivatives, at the first vision level, and learned filters at the upper levels. These filters are able to compute highly informative points or zones on natural scenes, such as: vertices, vectors, lips, eyes, mouth etc.... These image analysis tools are giving Neuro localization output information. They simulate the low level attractiveness of the Neuro-Visual system in the brain: Neurons can be considered as hardware filters, as they perform sum of products.
Moreover, such biological models can perform the perceptive grouping process, which merge similar textured regions, more efficiently that classic segmentation algorithms.

These information can give key elements for:

a) Finding corresponding points on epipolar lines: and achieve better Z depth estimation.
b) Helping a meshing tool for generating an optimized mesh.
c) Help the rendering algorithm to remove unnecessary details of the scene.

2 – MESH SIMPLIFICATION:

Biological models can be used also for reducing mesh complexity, according to the degree of importance of each picture area.

3 – ENHANCED RENDERING TECHNIQUES:

High quality rendering techniques have now important limitations in terms of displayed polygons per frame. Low cost hardware cards achieve now 1 to 2 millions polygons per second. But large, complex databases, such as architecture, environment need more polygons.

But it is quite unnecessary to display such polygon number when the viewer is watching only a limited part of his environment. Moreover, we must adapt the polygon number according to the solid angle of each observed object.

But how can we select important and non important area? Visual models can bring a response to this problem.

4 – MERGING IMAGE ANALYSIS AND SYNTHESIS:

A new direct bridge could be created between image analysis and synthesis, (as in MPEG4-SNHC goals), by the use of filter-based data structures instead of splines/polygons. Such data base could facilitate the gap between 3D objects and image processing elements. (The terms used in MPEG4/SNHC is VOP for video object planes: It's the first time that an image element other than contour, edge, etc., is considered as an object.
Possibility of automatic symbolization, as a tool for bridging the gap between 3D analysis and synthesis thanks to the Cortical Transform.

*Significant results have also been achieved in very low bit rate Coding (at 30 Kbits/s and less), with more acceptable aspects which fits with visual System requirements. Moreover, neural tools seem to be generic and well adapted to the various input styles of TV sequences, thanks to their learning capabilities & architecture dedicated to all kinds of 2D & 3D Analysis-Synthesis processing.*

***Keywords:*** 3D Image based Modeling and Digitizing, 3D image analysis, Neuro-focalisation and attractiveness, biological hardware filters, Neural Networks, 3D Image Analysis & Synthesis, Visual Systems, Simulation, Compression,MPEG4 hybrid coding.

## 1. IMAGE BASED DIGITIZING

Modeling 3D shapes is now a boring and expensive task, which don't help 3D development as a widespread technique.

Our goal is to mesh a landscape as simply as a camera or a scanner. Recently, camera makers are launching little digitizing laser based machines, whose output clouds of points and polygons. But 3D definition is very poor, and they need a laser beam.

Other system are using stereoscopic pictures, but only for little size objects. Moreover, they can't optimize the polygon mesh: For a cube as for a detailed human sculpture, the result is a huge polygon number. At the contrary, a graphist would have designed only a cube.

This is also the key problem for the mesh simplification: US academic people consider the 3D data input 3D data base "as is". But whether objects are digitized with a laser, and in this case, one can simplify drastically, because polygons are redundant, or objects have been modeled by a 3D modeling software, with a good graphist.

In the last case, mesh simplification is catastrophic!

The proposed technique is based on the ROI (Regions of Interest) algorithms: Such algorithms, based on vision models (See SPIE conference) are able to extract and quantize informative zones in the image. Such zones are depending of the filter size level. At different levels, one can obtain information such as: vertices, key elements i.e.: lips, mouth, etc.…

These areas are stable, after rotations or translations.

Therefore, disparity algorithms can extract corresponding points from several views. Moreover, once the camera position is computed from 8 or more corresponding points, such algorithm can help finding disparity for all pixels. Then, Zdepth extraction is more robust, and meshing can be done by Delaunay or Voronoï triangulation.

## 2. MESH SIMPLIFICATION

Neural based focus of attention algorithms can help mesh simplification because they can reproduce the cortical scheme, and fit with human requirements:

In the brain, we have two computing processes:

1. A bottom up process, which analyze images.
2. A top down process, which is a rendering process.

If we reproduce this scheme, it is possible to:

- Render a 3D scene , from a given point of view.
- Analyze the rendered image, using a multi-layer multi filter neural based filtering system which output areas of interest. Quantized responses from informative areas can be used for mesh refinement/simplification.

At this stage, it is possible to refine or simplify the geometric complexity, (Using simple geometric simplification algorithms such as Hoppe, Rossignac or Schroeder, controlled by informative zones ) according to the attention or area of interest filter activities. The process can then be reproduced, each time the viewer's position is moving.

Moreover, learning techniques could be used for object topology extraction.

## 3. ENHANCED RENDERING TECHNIQUES

Flight simulator were for a long time the pioneering domain for computer graphics enthusiasts. Now these techniques are currently used for industrial CAD design, Environmental simulation, Art, movies, etc.…

But some specific techniques currently used for real time simulation are not yet implemented on computer graphics PC's:

Object extraction is such a typical technique used for reduced polygon loads on 3D machines.

Flight simulator (For which I worked during 12 years) are dealing with huge 3D data bases (mainly 3D landmass terrain models), where polygon count reaches billions. It is mandatory to eliminate an important number of polygons, before loading the graphics card, when they are obviously outside the vision pyramid.

For this reason, if in-board hardware is of course able to clip each polygon, in fact this task is done for ALL polygons, and don't use any data base structure: Once one object is send to the graphics card, all his polygons are clipped.

Obviously, it would be more efficient to remove the entire object from the list being send to the card. Such process is called the extraction process, and can be done on standard PC's.

ROI computations can guided extraction process for selecting important parts of objects.

In our Viewer_Texture, we have 2 modules:

A VRML1 and VRML2 Viewer and process pre-computed Levels of Details (LOD)

A simplification module, which can reduce polygon number on any object: This module will receive the ROI algorithm for simplification optimizing.

Other refinements will also be added:

- Compute the “boxing process” i.e. in what part (box) of the scene the viewer is located.

- Display only potentially visible objects in the viewing pyramid.

- Redraw more often near objects, and less often far objects.

- Select adapted textures, according to the viewer’s position.

- Compute the texture Laplacian Pyramid for fitting with object/grid definition.

## 4. MERGING IMAGE ANALYSIS AND SYNTHESIS

Image Synthesis is generally based on shape representations using splines, nurbs, etc... at the upper level, and polygons at the lower level, very convenient for hardware acceleration.

Images analysis tools are mainly based on:

- Multi-level filtering process, using several algorithms such as sub-bands, wavelets,...

- Segmentation rules

- Motion estimation based on various block matching techniques.

These two kinds of representations and tools are very far from each other. But in the brain, we are storing obviously shapes and object representations, which can be manipulated as well as with image synthesis techniques, but these shapes are stored in the brain thanks to 2D & 3D analysis techniques.

The main process of the brain is the research of semantic abstraction, based on stored difference research.

This task is performed by a set of specialized cortical areas, such as MT, VT, ..., which separate different features for semantic extraction. A computational model of these areas has been designed by Professor Burnod in 1987.

This model is based on the biological concept of the cortical column paradigm in the visual area. Our simulator extracts 2D & 3D objects (1) shapes and movements by using the properties of hypercolumns (2) within the visual cortex, for spatio-temporal pyramidal filtering, learning, and performs inter and intra-cooperation between these simulated hypercolumns.

The simulation process has 4 abstraction levels for Analysis and Synthesis: Pixels, Zones, Objects and Labels. Final Synthesis (or reconstruction) is processed by reverse filtering, using non-orthogonal basis filters. Substantial upgrades in terms of compression ratios have been estimated using this algorithm as a whole, or partially, with integrated VLSI.

Current synthesis techniques are currently using shape representations, such as polygons, cylinders, splines, and so on. In the brain, we don't use such kind of representation, but a multiple pyramidal filter set, processed by elementary neurons:

basically, each neuron equals a filter, equals a convolution, equals a rule (as defined in IA techniques). The weights (coefficients) each neuron is using are the basis function of the filter; We can interpret the neural function in such term as:

Is the basis function $F_n$ I apply fits or not with the pattern vector Iv I receive in my synapses in input ?

Which is performed by:

Output = Sigmoid Function (SIGMA ($F_n$*$I_v$))

The neuron fires if the Output energy is high, and remains inactivated otherwise. At present new generic tools can be derived from human vision, image synthesis, artificial intelligence and neural nets. Each technique has advantages and limitations and they have to be combined in order to get better performance and generality. It is also interesting to note that efficient architecture is often close to the structure of the visual system associating complementary processes, such as scanning and focusing on points of interest, separation between a What and a Where pathway behind a natural scene, learning from examples, etc.. etc. ..

Recent neurobiological researches are leading us to new computational models of the human visual system. The cortical model has been presented several times, mainly in the SPIE/IS&T conferences in San Jose (CA, USA: see references).

This models is based on the cortical visual areas architecture and can be used for image analysis, 3D digitization, synthesis and compression

with 3 main advantages:

## 4.1 *The Human Visual System :* a model and a goal for artificial vision

The biological Visual system provides a general framework with outstanding results, largely better than our best artificial systems: It first enables us to interpret the effect of existing tools (image processing, neural nets) in relation with human visual processing : it is then possible to see how these tools can be combined to analyze pictures that will be perceptually acceptable and can fit with customer requirements.

## 4.2 Combined Analysis and Synthesis

CCETT has developed an important background in the field of 3D Image Synthesis, and designed in 1985 a 3D real time Image Synthesis System, which was called CUBI 7, dedicated to broadcast Production, CAD-CAM and Simulation. This machine has been industrialized in Europe, largely leading before SGI machines in 1985-86, and a European Project including Tübingen and Sussex University partnership issued a 3D graphics PC card and a 3D chip, which has now similar performance

as the best PC cards in the US.

At present, the major challenge in Computer Graphics is no longer the rendering process, but the 3D input digitizing process. Several techniques are now currently used:

- 3D modeling systems
- 3D Lasers
- Multi-views analysis systems

The third one seems to be very promising, but at present they need always manual interactive operations which are quite painful for a human operator. 3D Virtual reality will really emerge only when automated 3D digitizing tools will be commercially available.

In the other hand, Neural models of the Visual Cortex provide new tools for image analysis and synthesis, and more generally for scene and movement understanding; We have been developing an Analysis-Synthesis system, based on the same filtering process for the feed forward process (Analysis in input) and synthesis process ( for the output).

## 4.3 Hierarchy of Semantic Levels

Many image analysis systems operate at the pixel level, where the information data flow is huge. Classical neural nets associate only a large set of elementary neuron, with the same repetitive architecture: The result has poor training capacities.

On the contrary, the brain has a specialized, adapted architecture, which separates the visual process in elementary tasks, leading to less data, but from higher level, which is good for compression, and scene semantic understanding/digitizing. (3D data structures are no more than a semantic, compact representation of the real world).

Since the different types of low and high level visual processes are performed in the brain by the same type of neuronal circuit, the cortical column, we can propose principles for hardware architectures associating extraction capacities of image processing

tools and learning capacities of neural nets; furthermore such architecture could perform in parallel at different rates the main tasks unresolved by our computers:

- Object shape identification.
- Object movement
- 3D digitizing
- Compression with different levels of analysis
- Synthesis (Rendering).

## 4.4 Basic Modeling principles of the human visual system: a cortical architecture

### 4.4.1 Level: Pixels

The first step processes pixels : neuronal operators provide a set of complementary filters which extract specific features from the image: oriented and contrasted lines, end of lines, oriented movements, color contrast..

### 4.4.2 Level: Zones

The second step processes 2D1/2 zones: neuronal operators extract characteristics of local surfaces as homogeneous zones of pixels : extension, orientation of the surface in 3D space, colors and textures.

### 4.4.3 Level: Objects

The third step processes objects in a viewer-dependent reference axis: neural operators form clusters which code the main different aspects and views of a 3D object (in observer centered coordinates).

### 4.4.4 Level: Labels

The fourth step processes viewer-independent prototypes and labels: independent in size, orientation, and position of the retinal image. They also store the relations between prototypes (as semantic graphs).

This is the reason why we can introduce now the FILTER REPRESENTATION instead of the SHAPE REPRESENTATION classically used by image Synthesis techniques.

These 4 step levels cooperate through feed forward connections to perform the analysis of the image, from pixels to labels, and through feedback connections to perform a "synthesis" of the image, from labels to pixels.

### 4.4.5 THE SYNTHESIS (Reconstruction) PROCESS

Classic coding schemes are generally using orthogonal basis filter functions, with fixed coefficients, which are very convenient for a perfect reconstruction. However, these basis are rigid, image independent, and in consequence poorly adapted to particular image characteristics. In our Cortical Transform paradigm, we use the same filter basis functions for analysis AND synthesis, even if this filter basis is not orthogonal. In consequence, we can choose filters whose neural activity is strong, both for analysis and for synthesis: It is obvious that the stronger the neural activity, the easier the corresponding filter will be adapted.

Let I(x,y) the luminance value be at x,y; Gi(x,y) a set of non orthogonal basis functions, for which each neural activity is ai. We have at scale k:

ak = Si I(x,y)*Gik(x,y)

If the basis functions are orthogonal, we can reconstruct exactly I(x,y) using:

I(x,y) = Sk ak*Gik(x,y)

In our model, functions are not orthogonal, and ak are adjusted using an iterative neural scheme, minimizing the error:

$|e|^2 = | I(x,y) - S\ ak*Gik\ (x,y)|^2$

Such a process can be generalized at each abstraction level, making it possible for the cooperation between analysis and synthesis. In particular, 3D objects can be digitized using a set of Neuron-Filters, synthesis being performed by such a reverse filtering process.

Using the feature extraction we have described above for image analysis and compression, it is possible to propose a complete compression-decompression scheme based on this paradigm. This system is composed of two Analysis-Synthesis modules, one for the coder, and the second for the decoder.

## 5. RESULTS

### 5.1 Image based digitizing

We have been implementing a full ROI algorithm based on cortical models (see results on annex 1), at different filter resolution. Our system can detect interesting zones or areas, such as mouths, lips, or even edge corners, without pre-training.

This system is currently under development before integration in a complete image based 3D digitizing system.

### 5.2 Mesh simplification

Similar ROI algorithms will be used for helping automatic mesh simplification.

### 5.3 Enhanced rendering techniques

A complete VRML1 and 2 Viewer has been developed (see images in annex 1), including:

- LOD (level of detail) automatic processing.
- Textures computation.
- Scientific tests have been made on real ancient architectures or objects:

  Mont St Michel – St Germain Church Pipe Organ (17th Century).

### 5.4 Merging image analysis and synthesis

Is a target development planned for further researches: It gives an interesting research avenue for merging these two techniques.

## 5.5 Compression

Non orthogonal multi-semantic levels basis functions have been investigated: In our last presentation (MPEG 4 Paris) we cut by 2 the transformed entropy;

For still image compression, a compression ratio of 130 to 190 was achieved with better visual quality than JPEG.( Lena: 0.0228 bit/pixels (compression ratio:130 ) Susie: 0.0158 bit/pixel (compression ratio: 190).

For animated images: Perceptive grouping & attention zones, and more generally feature extraction has been implemented using spatio-temporal filtering. Test sequences have been used - at 30 kBits/s) – (SPIE 94 & 95).

# 6. CONCLUSION

Neural based vision models provide outstanding tools for the most complex problems we have in 3D automatic digitizing, level of detail processing, and semantic real world computing. All implementations are not yet achieved, but significant partial results are the proof that such technical skill are key elements for Computer Graphics and 3D Image analysis.

# 7. REFERENCES


[1]Burnod, Y. "An Adaptive neural network: the Cerebral Cortex." Book: 400 pps. Masson; Prentice Hall 1988.

[2]Leray, P. "A 3D Synthetic Imagery Generator in real time" presented at the: Image Generation/display Conference II Air Force Human Resources Laboratory, Phoenix, AZ USA pp 78-89 1981.

[3]Leray P. "A 3D animation system Datastructures for Raster Graphics" Data Structures for Raster Graphics Springer-Verlag pp 165 - 171 1985 .

[4]Leray P. "Towards a Z-Buffer and Ray-Tracing Multimode System based on Parallel Architecture and VLSI chips", Advances in Computer Graphics Hardware 1 Springer Verlag pp 141 - 145, 1987.

[5]Leray P. "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.

[6]Leray P. Burnod Y. : "CUBICORT: A simulation of a multicolumn model for 3D Image Analysis, understanding &compression for Digital TV, HDTV & Multimedia" 1993- Laurence Erlbaum Associates, Publishers - Hillsdale, New Jersey

[7]IWANNT: International Workshop on Applications of Neural Networks to Telecommunications.

[8]Leray P. “New advances in Neuro -Visual Simulation and Symbolic extraction for Real World Computing, 3D Image Analysis and 3D object Digitization“

[9]Race Project MAVT Seminar: Rennes - France 94.

[10] Leray P. F. Guyot, P. Marchal, Y. Burnod "Simulation of the Visual System for 3D Image Analysis, & very low bit rate compression for Digital TV, HDTV & Multimedia" 1994- SPIE-IS&T San Jose USA January 1995

[11]Eurographics Hardware Workshop 1996: New advances in Neuro -Visual Simulation and Symbolic extraction for Real World Computing, 3D Image Analysis and 3D object Digitization.


## Author:


Pascal Leray

France TELECOM/CNET/DIH

Centre National d’Etudes des Télécommunications

CCETT Centre Commun d'Etude de Telediffusion et de Telecommunication

Rue du Clos Courtel, BP 59 - 35510 CESSON-SEVIGNE

(a FRANCE TELECOM Research Center)

**Email:**

pascal.leray@cnet.francetelecom.fr

perso.wanadoo.fr/pascal.leray