

SCIENTIFIC AND TECHNICAL REPORTS SUMMARY

FINAL REPORT

CONTRACT DASG60-95-C-0080

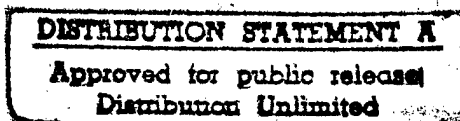
Effective Date of Contract: April 25, 1995

Expiration Date of Contract: October 25, 1995

Reporting Period: End of Contract

Principal Investigator and Phone Number: Dr. Thomas Tsao, (301) 680-0443

Short Title of Work: Neural Geometric Engine Feasibility Study



19951107 035

DTIC QUALITY INSPECTED 8

CompuSensor Technology Corporation

Introduction

The goal of this project is to research and develop a neural geometric engine for rapidly determining geometric relations between parts of a scene from sensor images. The subject of building a spatio-geometric and kinetic model of the scene from images was considered "image understanding" or "early vision" in artificial intelligence research.

The approach we have taken to spatio-geometric modeling of the scene is a **smart sensor** approach. It is fundamentally different from the current art. The novel neural computing system is based on Lie group model of neural processing in primate's visual cortex.

Termed "**information processing**" approach to vision by David Marr, the pioneer of computational vision research, the current art of early vision is build upon the concept that the spatio-geometric information can be extracted by processing the image data, and the process can be formed as a computer algorithm.

While the term "information processing approach" sounds very general, it does lead to a specific method of algorithm design. Particularly, it was suggested that in order to determine the changes (motion, binocular disparity, geometric distortion) in images and to further infer the scene geometry and motion, or register images, the first step should be to determine how a point on the image plane is moved to another place. It was further suggested that a process of **feature detection** followed by **feature matching** will do the job. All the spatio-geometric information are considered directly or indirectly derived from feature matching. It appears to be a very natural and very common sense approach to follow except for a little difficulty in its logic.

In order to measure geometric changes from the images, the computation must anchored to some recognizable place holders, the image features. If a feature is a dot type place holder, it provides no cue for matching: One such place holder does not distinguish itself from the others. If it is a patch of image, itself will subject to changes. In order to match patches, the changes must be compensated while the very changes are to be computed! The current art of getting out of the bad loop is some trial and error, some heuristic control, some tolerance of error, some constraints, some middle ground taking, etc. Some of these strategies are of ad hoc nature, others are with deep thinking. All kinds mixtures of these ingredients are flourished in a beautiful garden of image understanding with tens of thousands technical papers published there.

The "information processing" approach to visual perception was criticized by J. J. Gibson, one of the most influential psychologist on visual perception research in America. According to Gibson, the spatio-geometric relation is contained in the visual stimulus, and can be **directly picked up** by vision system. It is a **smart sensor** approach. The assertion is, the vision system does not manipulate the image data to "compute" the geometric information, but simply pick it up by the direct response to the stimulus. Marr's criticism to it was that the smart sensor approach grossly underestimated the complexity of visual information processing. While the

<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
des	
Dist	Avail and/or Special
A-1	

information processing approach was supported by the firm ground of modern computer technology, Gibson's approach was supported by only a firm philosophical conviction that vision system is an instrument for animal to adapt to its environment. For that reason, the nature of visual perception must be a kind of direct response to visual stimulus justifiable for animal's adaptation to the environment. Lacking of computational theory and practice, Gibson's theory was moved to the back stage, and was regarded as a philosophy.

Two big problems caused by finding the feature matchings are the **combinatorial complexity** and **uncertainty**. They make the accurate and robust geometric modeling of the scene virtually impossible, and prevented images from being used as effective sensor means. For example, it is easy to get binocular stereo images. However, to date no computer based system uses binocular image pair to generate 3-D surfaces. Instead, 3-D images are mainly generated by active sensors, such as laser range finder, structured light 3-D imaging system, etc. For the same reason, image registering, image fusion, object recognition, object motion computation, all suffered same problems of combinatorial complexity and uncertainty.

There are persistent efforts of developing new computer architecture to overcome the above mentioned problems, and to make the collected image data more useful. These efforts use parallel processing, faster processors, and other techniques to increase the speed of computers. Still based on the basic method of **feature detection** and **feature matching**, these approaches are brute force by nature. The success of brute force approach to early vision problems are very limited.

It was until 1980's, that neurobiologist started paying great attention to the dynamical property of the receptive fields of cortical cells. It was observed that cells in primate's visual cortex can maintain a stable response to an object in motion by adaptively shifting and warping their receptive fields. The dynamics of the cortical neuron reveals the secret of how the smart sensor is build. The process can be modelled using Lie group method. This leads to another theory of early vision, a theory of how the brain can adapt to an environment with motion and spatial disparity to maintain an invariant representation of the object of concern, and to obtain a spatio-geometric model of the environment.

The cortical neurons with dynamical receptive fields thus perform the function of a smart sensor capable of directly picking up the image geometric transform information, as Gibson had suggested. The smart sensor can be implemented using analog VLSI technology to mimic the analog process in primate's visual cortex. It also can be implemented as a digital system using fixed amount DSP chips to cope with the required computing power. In either of the implementations, the resulted system is a simulator of the particular neural circuit in primate's visual cortex, and is called a neural geometric engine. In either implementation the architecture of neural geometric engine is derived from the Lie group model of primate's visual cortical process.

This report summarizes our phase I effort in research and develop the neural geometric engine.

1. Task Objectives Achieved

Three objectives have been achieved through phase I research: (1) verify the validity and robustness of the basic computational structure of the neural geometric engine; (2) outline the architecture of the computing system, including the information coding method, the selection of computational primitives, the Lie group processor, and the organization of the system; and (3) confirm the availability of hardware technologies suitable for digital or analog implementation of the neural geometric engine.

The proposed neural computing system is different from neural network models for pattern recognition. The neural networks for pattern recognition are based on various models of associative memory. The core parts in these neural network models are the learning algorithms by which the networks can build up associative memory for carrying out particular pattern recognition tasks.

The neural geometric engine is a perceptual engine. Its task is **not to build an associative memory** through a learning process, but to **build up a geometric-kinetic model** of the scene in responding to image input in real-time. The spatio-geometric perception of a scene is accomplished by several levels of visual processing. The first level process is to determine the local affine geometric transformations in image sequence or in binocular image pair. It is a true **leap** from which the brain starts perceiving its environment in terms of *geometric parameters* while originally it only has *sensor signals*.

Instead of functioning as associative memory, or feature detectors, a substantial part of primate's primal visual cortex has the function of a dynamical coordinate system for visual stimulus. They are organized in hypercolumns consisting many orientation specific microcolumns. The receptive fields of these cells not only serves as **basis functions** for encoding the local oriented contrast of visual stimuli, but also can adaptively change in real-time to maintain stable percepts of objects in motion. These cells provide a **moving reference frame** for images. The moving reference frame is a smart sensor which responding to the transform of visual stimuli with its own transform.

The perceptual leap is achieved via a dynamical process facilitated by a neural circuit. The neural circuit, consists the neural computational elements for cortical representation of visual information, cortical coordinate affine transforming, feedback control, and Lie germs, is called a **Lie group processor**. The Lie group processor defines the basic computational structure of the Neural Geometric Engine. Our first objective was to verify the validity of this computational structure.

Computer experiments establish the feasibility of our new concept and method. Our results showed that for affine transforms up to 30 degrees rotation and 80% linear scale, the digital simulation with the Newton scheme converges in a few iterations with error less than 5%.

The phase I work outlines the architecture of the neural geometric engine and provides a theoretical foundation for the novel neural computing system:

- (a) In the neural geometric engine, visual information is represented as the measurement of image intensity by **linear receptive fields** which are modelled as various derivatives of Gaussian distribution functions, and the measurement of relative geometric deformations between parts of different images by Lie germs.
- (b) The basic computation of the system is a nonlinear dynamical process with a minimum energy state. The process involves the operations of linear cell receptive fields and the operations to transform these receptive fields in a feedback loop. This basic computation is supported by a physical circuit called **Lie group processor**.
- (c) The primitives for linear cell receptive field processes are multiplication and summation. For affine transforming the receptive fields functions, the system further includes exponential mapping as a primitive function. This is because the receptive fields take the Gaussian distribution function as the basic form of the spatial extension. In a word, we chose three computational primitives: **summation, multiplication, and exponentiation**. All of them can be implemented by fundamental physical phenomena of analog circuits.
- (d) The neural geometric engine is a hierarchical distributed information processing system which includes two levels of function: it first extracts the affine parameters of local image transform from images, and then computes from these parameters three dimensional motion and shape of objects.

The phase I study identifies several high end parallel computing systems and state of the art DSP chip technology for building a digital version of the neural geometric engine and achieving real-time or near real-time performance for several important applications. Phase I study also confirms the availability of analog VLSI technology for implementation of the neural geometric engine. Analog VLSI implementation makes possible particular military applications that require miniature size and very low energy consumption.

2. Technical Problems

Real-time determination of the spatio-geometric relation between parts of a scene from sensor images is the key to various autonomous systems. The problem appears in various places with different forms, such as automatic terrain recognition for robotics vehicles, automatic target recognition, sensor image fusion, stereo surface characterization, image motion compensation, etc.

Images collected by sensor systems mounted on moving platforms or from multiple sensors, and images of moving objects, are subject to geometric transformations. The parameter of image geometric transformation carries geometric and kinetic information of the environment, such as three dimensional structure of visible surfaces, and object or platform motion. In order to recognize objects in various poses, fuse sensory data collected from different sensors, a common problem is to reduce transformational differences of image data. Also, in practical situations, it often happens that real-time computation is required.

Thus the general problems in early vision are: (1) How to determine from image data the geometric parameters of the scene, and (2) How to determine it in real-time. Our approach to these problems is to build a neural geometric engine.

The following specific technical problems for building a neural geometric engine are those with regard to the architectural issues and the implementation issues:

- (1) Define a representation scheme for the visual information in this neural computing system;
- (2) Verify the fundamental computational structure of this neural computing system;
- (3) Define the computational primitive set of this neural computing system;
- (4) Define the organization of this neural computing system for the early vision process;
- (5) The approach of implementing this neural computing system with digital means; and
- (6) The approach of implementing this neural computing system with analog means.

To systematically resolve these technical problems requires extensive and specialized research and development effort involving areas of computational vision, mathematical modeling of biological visual cortex, neural computing theory, parallel and distributed digital computing method, DSP computing technology, and analog VLSI computing technology.

3. General Methodology

Phase I work is focussed on the feasibility of implementing the neural geometric engine with practical applications, and exploration of commercial potentials. The feasibility study includes verifying the validity of the fundamental computational structure, the survey of hardware technology suitable for implementing the engine, the collection of practical problems targeted for the neural geometric engine to solve, the experiments with examples of these problems. The phase I study therefore involves computational experiments, literature survey, visiting the potential users, collection of examples of practical problems and real data, experimenting with these application data.

1. Computational Experiments

The concept of extract geometric transform parameters from intensity images through a dynamical neural process of "receptive fields", modelled as cortical coordinates and Lie derivative operators, is new in mathematics as well as in computational vision and image processing. There is nothing similar to this work that we can borrow or get some guidance from. Whether the mathematically verified numeric procedure will work in actual computer experiments is a first question. Also we need to see how fast the algorithm will converge to a solution and how accurate will it be when it converges. Without answering these fundamental questions with actual computations, further research and development of the computational structure, the algorithm and architecture, as well as applications, will be baseless.

In phase I study, both real image data and synthetic images are used in the computational experiments. The advantage of using synthetic images is that the accuracy of the computation can be directly measured because the actual geometric transform of data are known exactly. Experiments with real images are necessary because real imagery are usually noisy, and with background clutter. A robust algorithm must be graciously degrade its performance as these noise and disturbances are presented. Also the accuracy should be recovered if data redundancy is plentiful.

2. Literature Survey

Behind the design of a digital computer is the whole knowledge body including the theory of digital computing (mathematical logic and algorithm, computability), theory of symbolic information coding (information theory), the methods of symbolic data structures and file organization, and designs of digital electronic hardware architecture, *etc.* There is no such well formed theoretical base and knowledge body available to date for designing a neural computing system.

However, some fundamental theoretical problems must be answered if our approach is not of *ad hoc* nature. For that purpose, extensive literature review and survey has been an essential part of work for defining the architecture of neural geometric engine.

Particularly, we have reviewed and surveyed articles regarding to the coding method of analog signals and visual information, articles on visual perception process and artificial vision, the recent results in biological study of visual cortex of primates, the recent technological development in analog VLSI computing, the recent DSP chip technology, the parallel and distributed digital computing, dynamical system theory, and mathematical modeling of neural learning process and neural computing in general.

The extensive literature review and survey has helped us to crystalize our concept of the geometric engine and the way of implementing it.

3. Collection of Application Problems and Examples

During the phase I research, a set of examples and problems are collected through contacting to potential users, having technical discussions with them, and taking their problems and examples for studying.

4. Technical Results

1. Verification of The Computational Structure of The Lie Group Processor

The process of adaptive change of receptive fields of neurons in response to the change of visual stimuli is a basic process of V1 area. It represents the function of the Lie group processor. It kills two most difficult problems in computer vision, the affine invariant feature extraction and the so-called "feature correspondent problem", in one shot.

The question can be set forth as this: given two actual image patches, one a transformed version of the other, can a machine determine the parameters without using traditional computer algorithm tricks, such as feature matching, trial and error, artificial intelligence heuristic, knowledge, etc., but **simply by the dynamics of a feedback circuit**? The departure from all other approaches and the start of neural geometric computation will be possible only if this test can be passed.

Related to the above question is: to what extent can the scheme determine the parameters of image transforms? Any realistic application will demand the computational scheme work in a range of parameters that has practical significance.

To answer these questions, a set of simulations have been done. The result confirms our conviction that the Lie group method will be a superior method for early vision processing.

A sequence of computer experiments were performed on a Pentium PC. In these experiments, all the Lie group parameters are initially set to zero (scale parameter set to 1). With Newton scheme, we found in many cases, the first iteration is able to get very close to the true transformation parameters, and thus substantially reduce the "energy".

Figure 1 shows a computer generated target pattern and its transformed version which is rotated 15 degrees and scaled by 0.8 in both dimensions. Figure 2 shows the geometric compensation process reducing the difference between these two patterns measured by the "energy" in a dynamical process. Figure 2(a) is of the gradient scheme, (b) is of the Newton scheme.

Figure 3 shows a computer generated target pattern and its transformed version of rotate 10 degrees and scale to 1.2 in both dimensions. Figure 4 shows the geometric compensation process reducing the difference between these two patterns measured by "energy" in a dynamical process. Figure 4 (a) is of the gradient scheme, (b) is of the Newton scheme.

Figure 5 shows a computer generated target pattern and its transformed version which is rotated

20 degrees and scaled by 0.9 in both dimensions. Figure 6 shows the geometric compensation process reducing the difference between these two patterns measured by the "energy" in a dynamical process. Figure 6 (a) is of the gradient scheme, (b) is of the Newton scheme.

Figure 7 shows a computer generated target pattern and its transformed version which is rotated by -15 degrees and scaled by 0.85 in both dimensions. Figure 8 shows the geometric compensation process reducing the difference between these two patterns measured by the "energy" in a dynamical process. Figure 8 (a) is of the gradient scheme, (b) is of the Newton scheme.

The results of the computational experiments in terms of geometric transform parameters been determined with the four synthetic image patterns in the dynamical processes are listed in the following table:

Pattern Class	Transform Parameters	Compensated in Gradient Process	Compensated in Newton Process
1	$\theta = 15^\circ$ $\sigma = 0.8$	$\theta = 15.0531^\circ$ $\sigma = 0.8139$	$\theta = 15.0611^\circ$ $\sigma = 0.8068$
2	$\theta = 10^\circ$ $\sigma = 1.2$	$\theta = 10.0460^\circ$ $\sigma = 1.2220$	$\theta = 10.0452^\circ$ $\sigma = 1.2168$
3	$\theta = 20^\circ$ $\sigma = 0.9$	$\theta = 19.8470^\circ$ $\sigma = 0.9182$	$\theta = 20.0714^\circ$ $\sigma = 0.9149$
4	$\theta = -15^\circ$ $\sigma = 0.85$	$\theta = -14.2633^\circ$ $\sigma = 0.8623$	$\theta = -15.1310^\circ$ $\sigma = 0.8579$

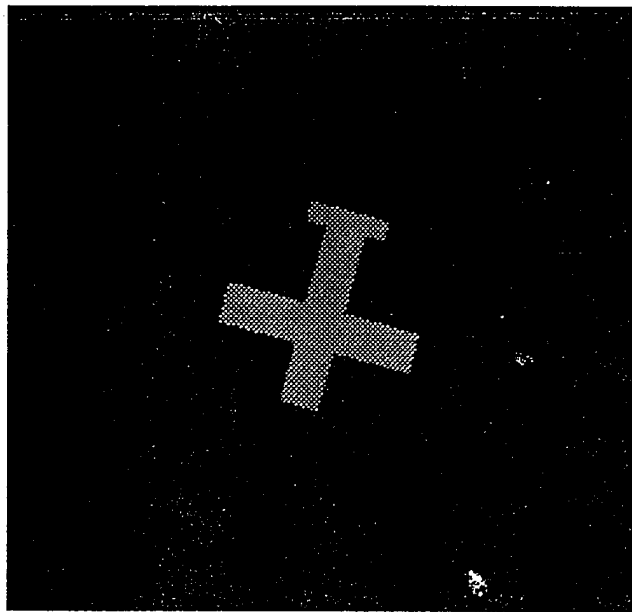
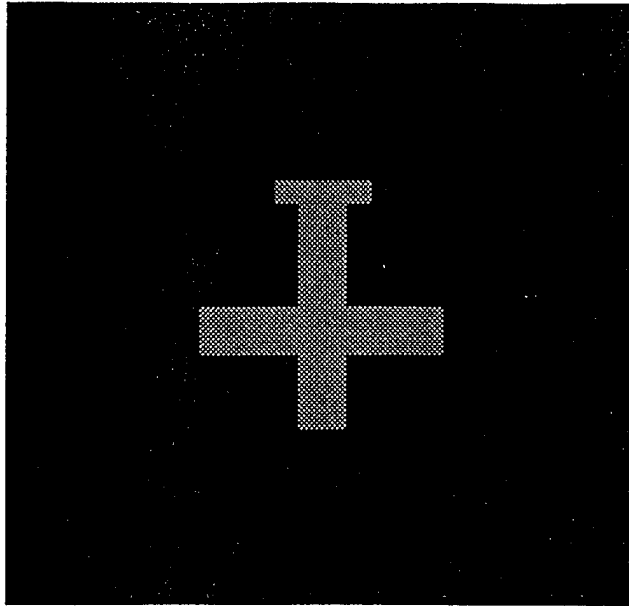


Figure 1. A computer generated target pattern and its transformed version of rotate 15 degrees and scale to 0.8 in both dimensions.

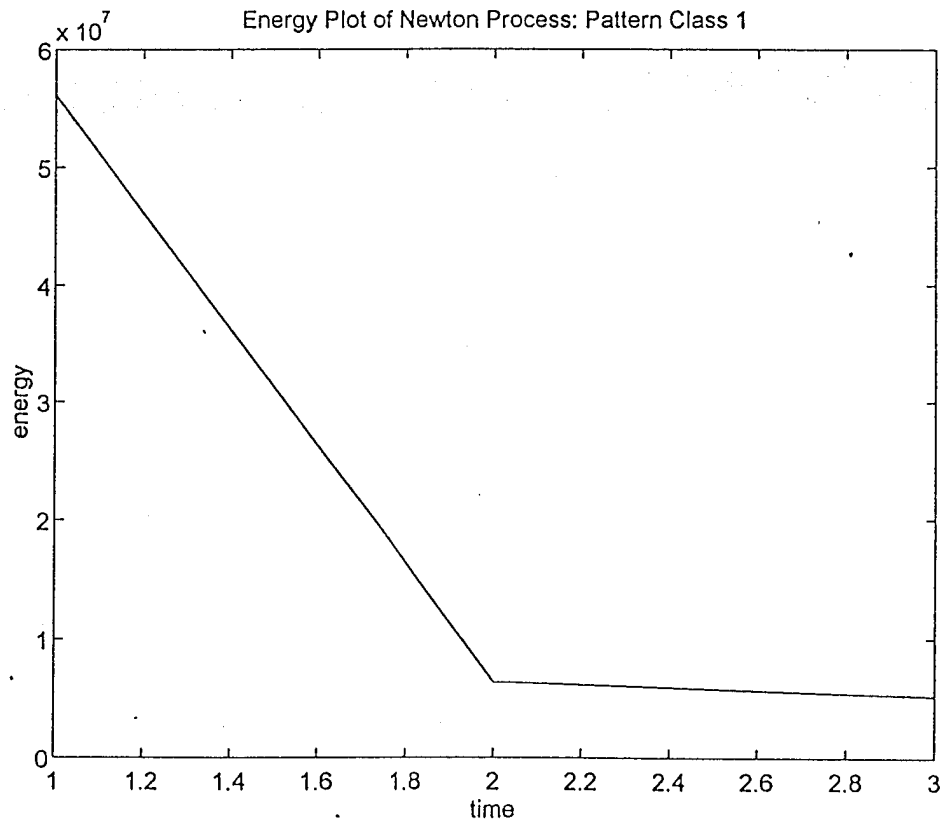
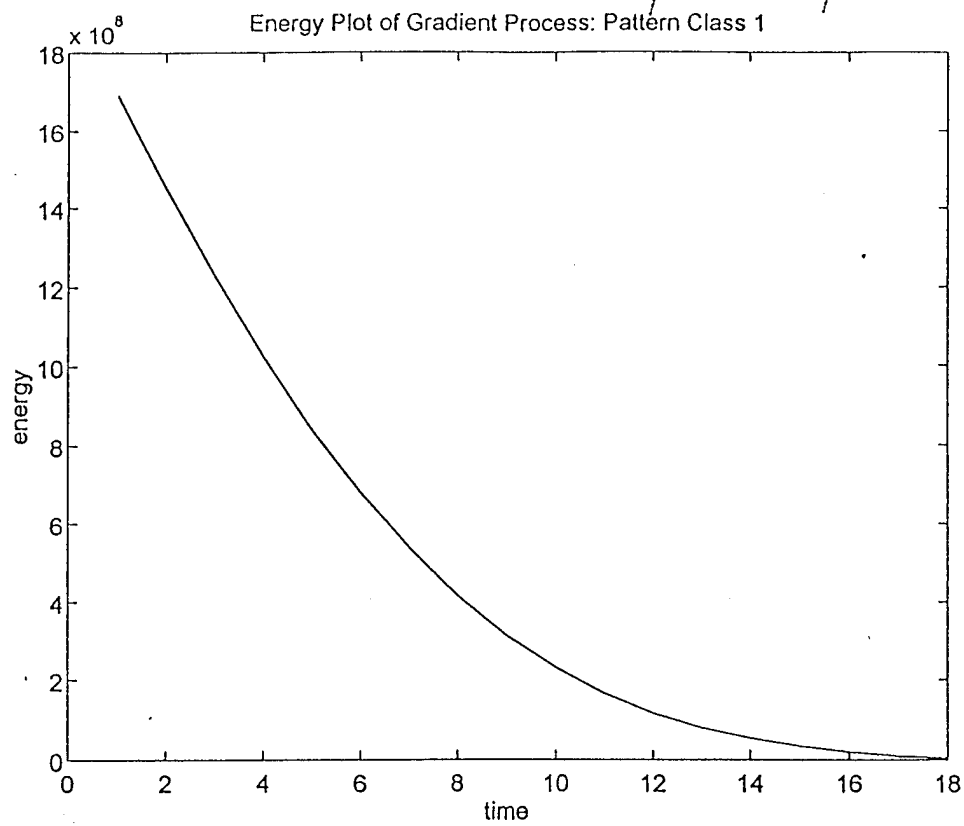


Figure 2. The geometric compensation process reducing the difference between these two patterns measured by "energy" in a dynamical process. (a) is of the gradient scheme, (b) is of the Newton scheme.

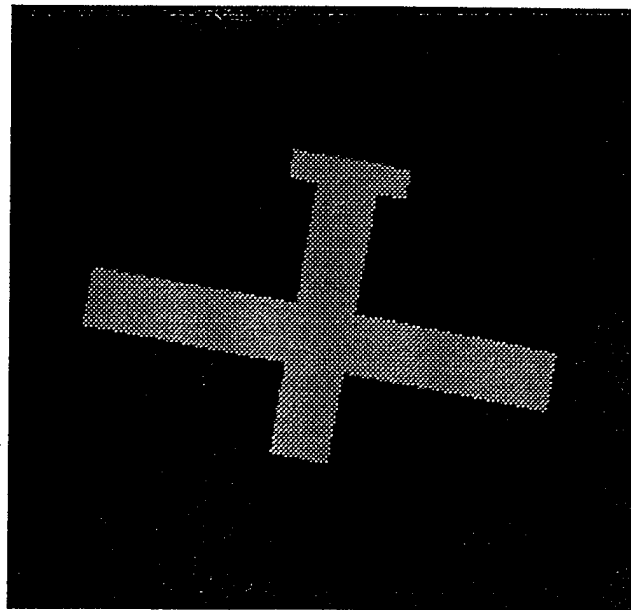
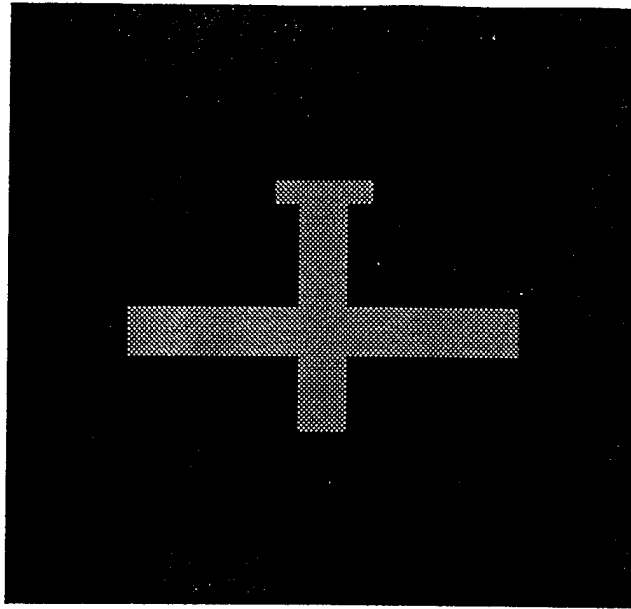


Figure 3. A computer generated target pattern and its transformed version of rotate 10 degrees and scale to 1.2 in both dimensions.

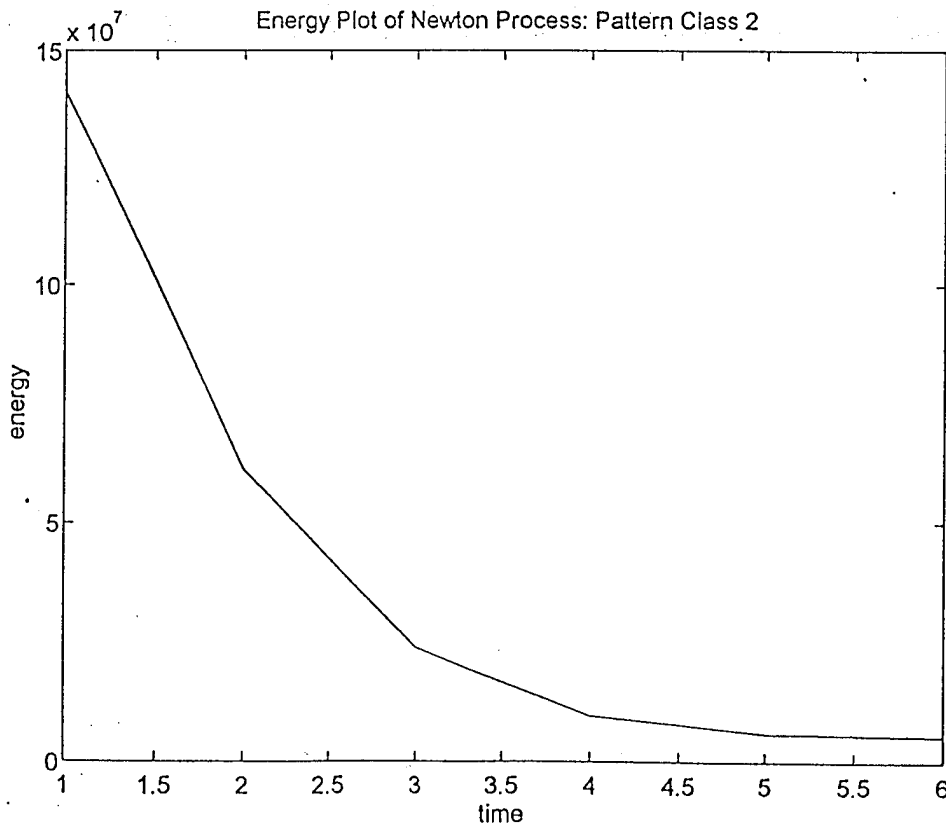
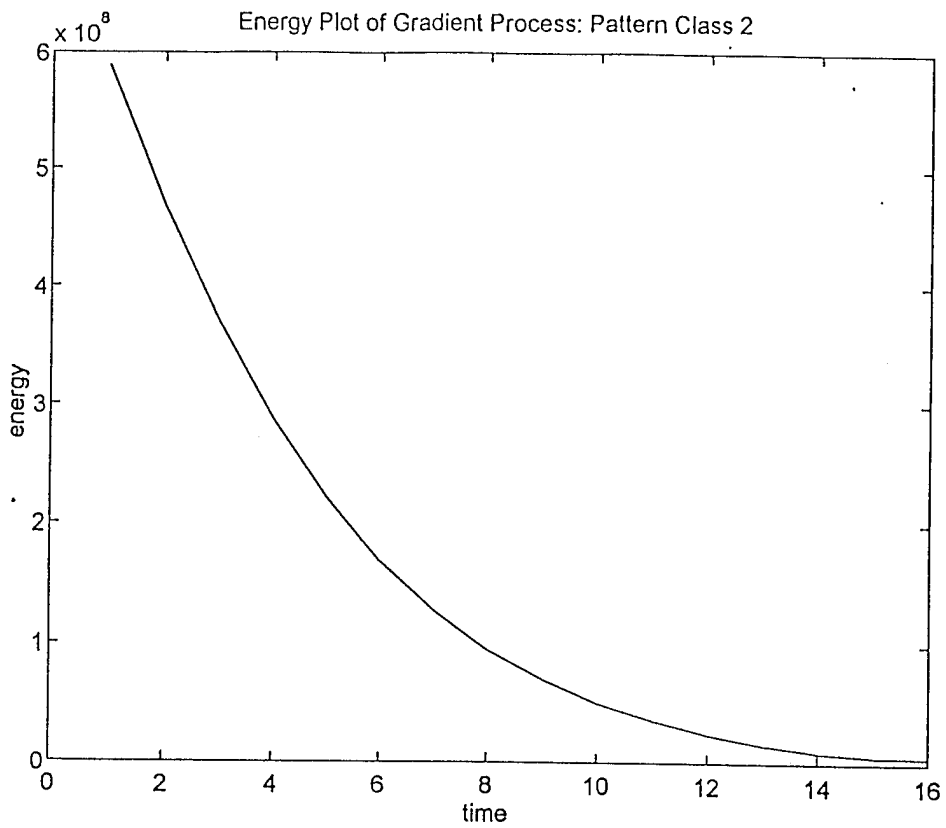


Figure 4.1. The geometric compensation process reducing the difference between these two patterns measured by "energy" in a dynamical process. (a) is of the gradient scheme, (b) is of the Newton scheme.

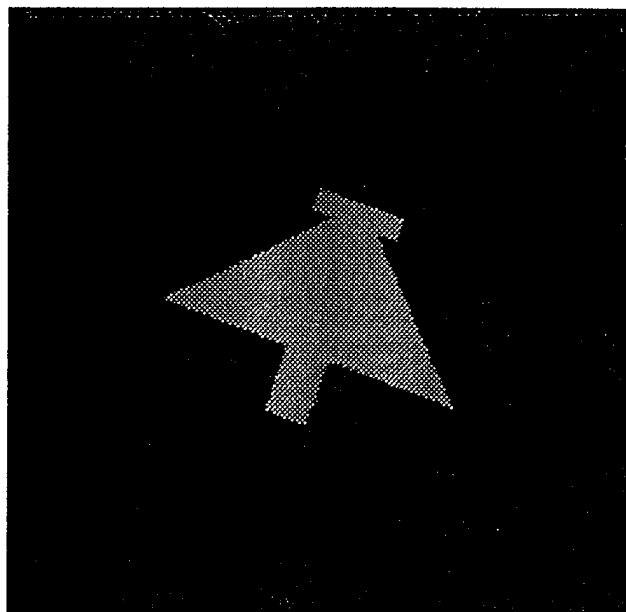
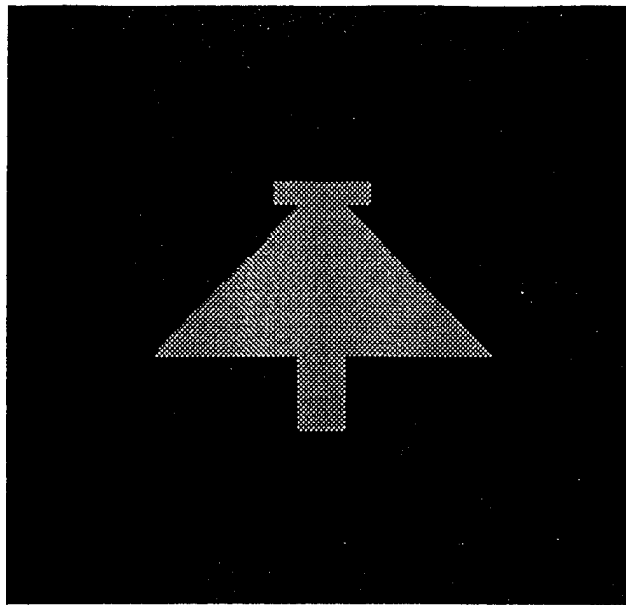


Figure 5. . A computer generated target pattern and its transformed version of rotate 20 degrees and scale to 0.9 in both dimensions.

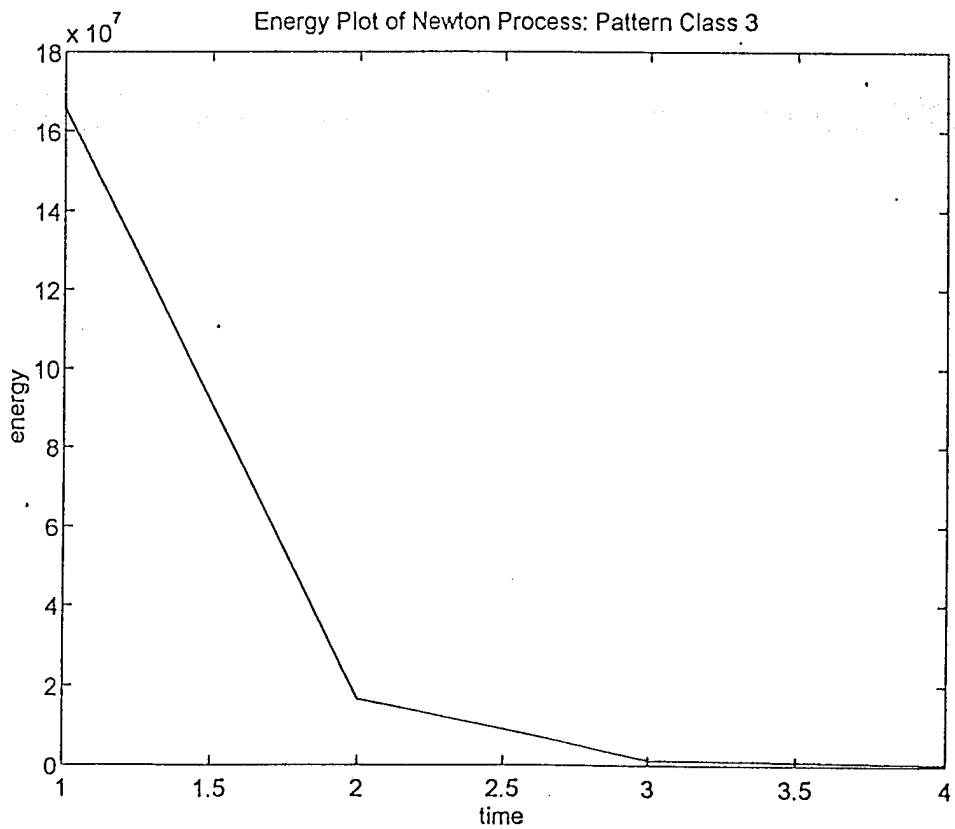
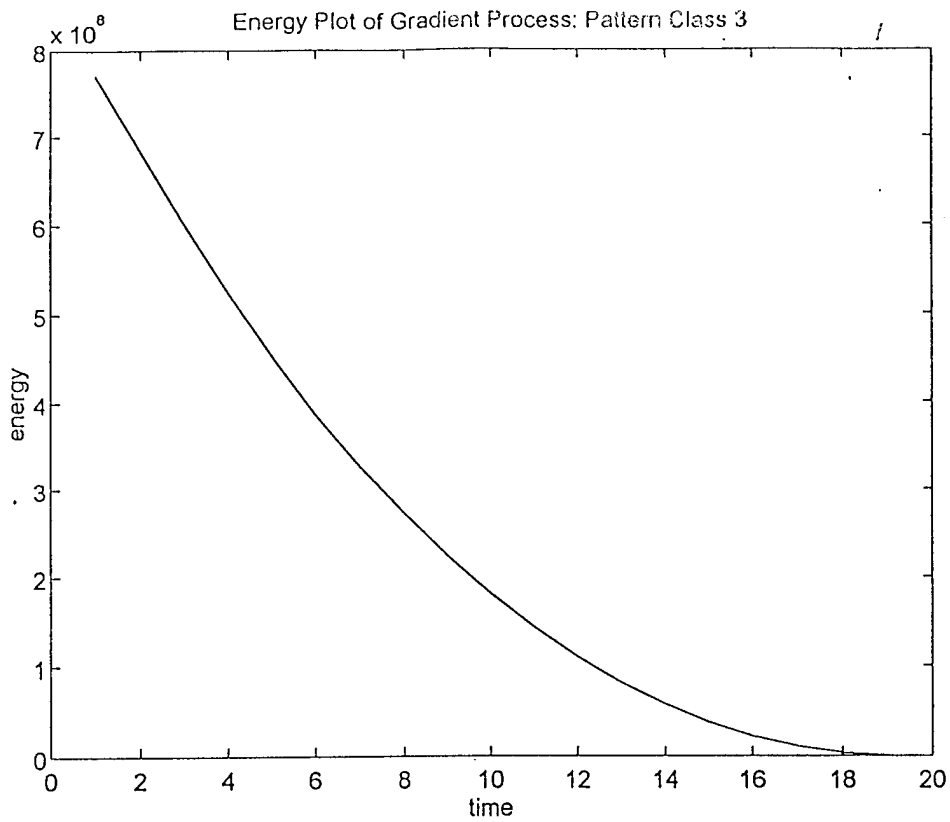


Figure 6.: The geometric compensation process reducing the difference between these two patterns measured by "energy" in a dynamical process. (a) is of the gradient scheme, (b) is of the Newton scheme.

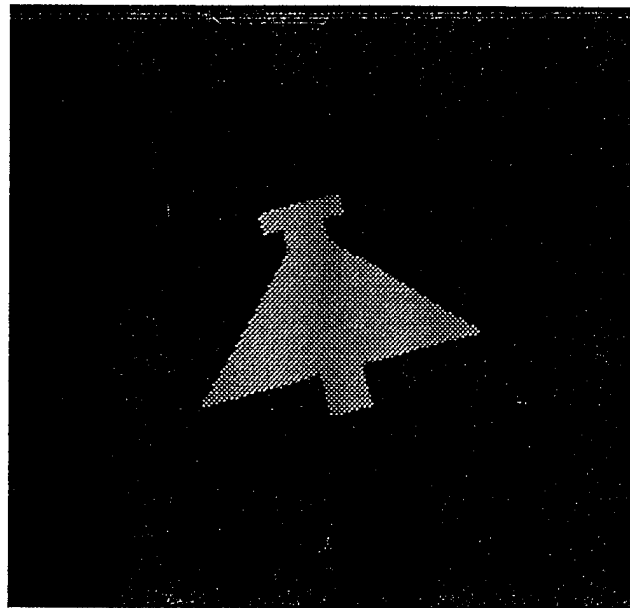
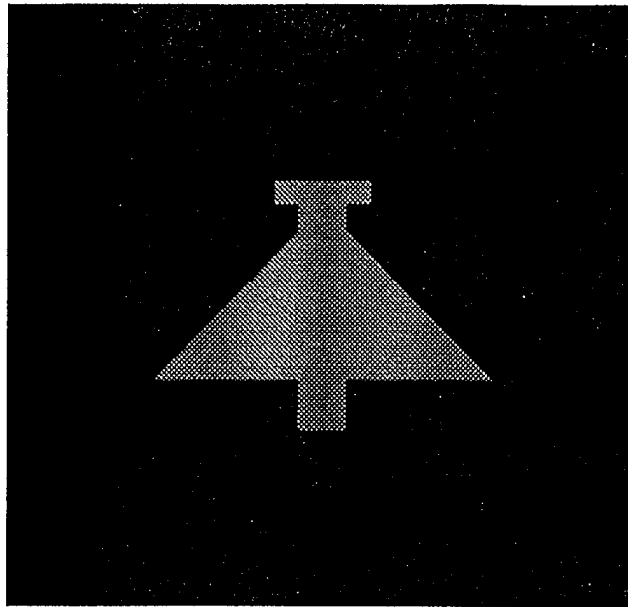


Figure 7. A computer generated target pattern and its transformed version of rotate -15 degrees and scale to 0.85 in both dimensions.

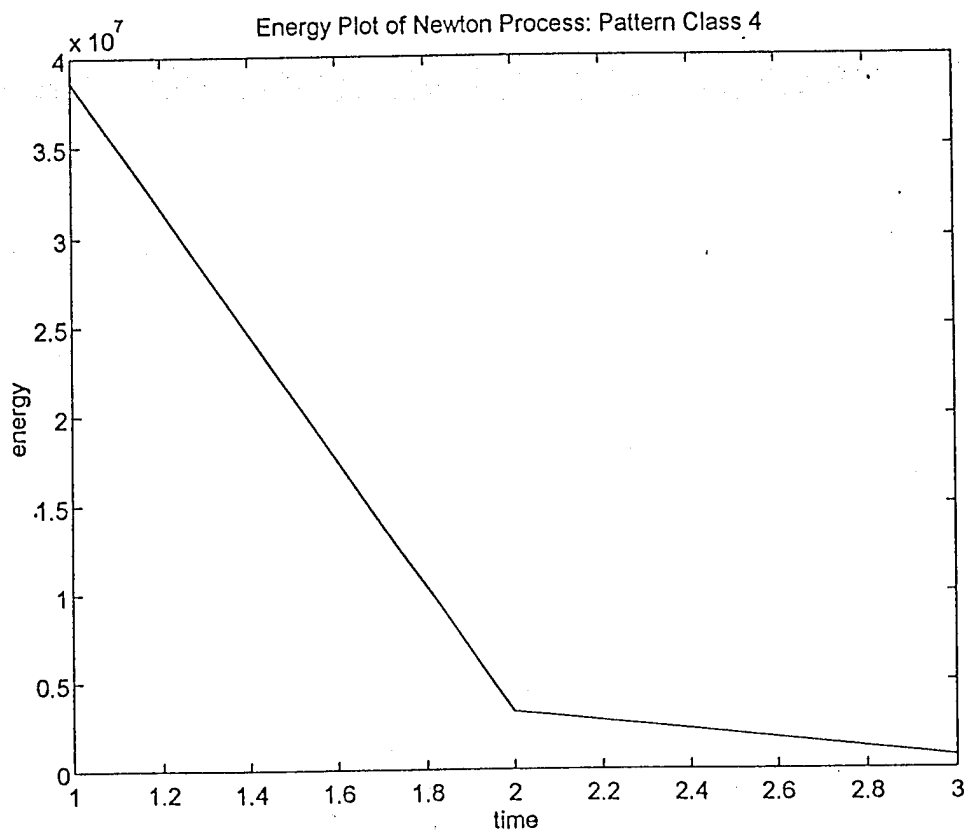
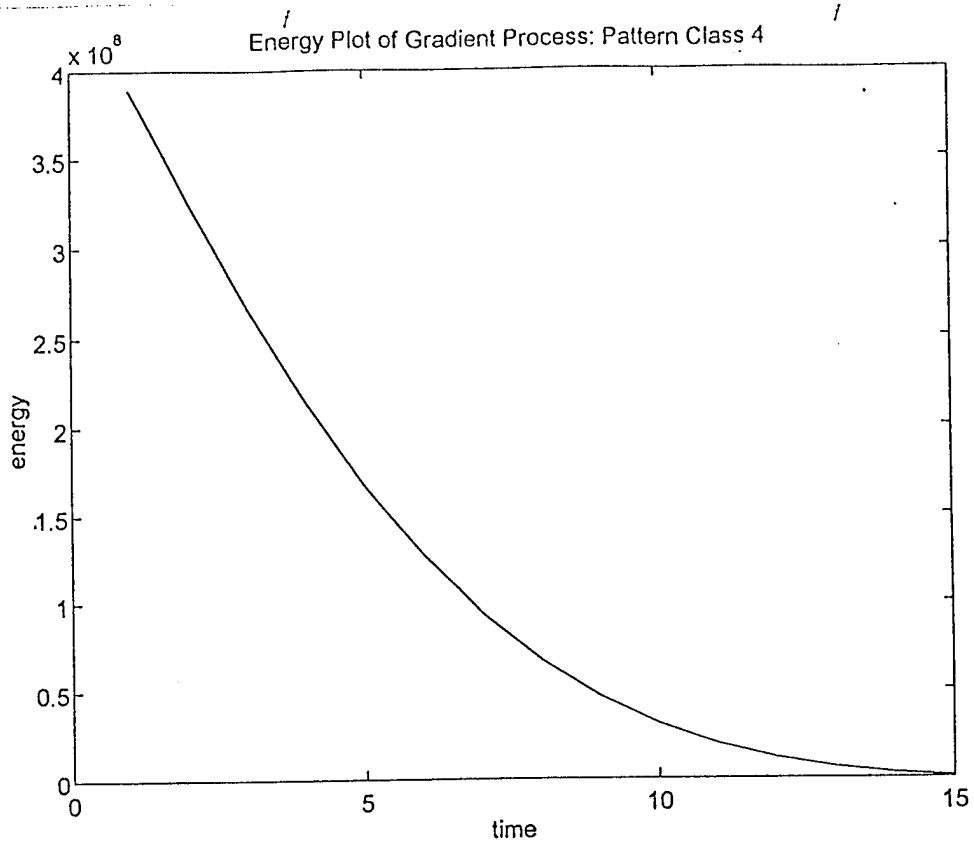


Figure 8. . The geometric compensation process reducing the difference between these two patterns measured by "energy" in a dynamical process. (a) is of the gradient scheme, (b) is of the Newton scheme.

2. The Architecture of The Neural Geometric Engine

1. *The Concept of The Neural Geometric Engine*

The neural geometric engine is a real-time computing system designed according to neural computation methods for extracting spatio-geometrical information from a scene.

The neural computation methods include (1) a method of *neural representation* of information, (2) a method of *neural processing* of information, (3) a set of *neural computational primitives*, and (4) a *neural organization* of information processes. There are fundamental differences between digital computing systems and neural computing systems.

In contrast to digital computer systems where information is represented by the absolute value of digital signals, in the brain, sensor information is represented by the relative value of analog signals.

Accordingly, in the neural geometric engine, visual information is represented by the measurements of intensity image by receptive fields which can be modelled as various spatial derivatives of Gaussian functions or Gabor functions and the measurements of relative geometric deformations between different image parts by Lie germ type hypercomplex cells.

In contrast to digital computer systems where information (represented as discrete symbols) is processed according to algorithms which should halt in a finite number of steps, in the brain, the sensor information (analog signal) is processed by nonlinear dynamical systems which yield definite results when they converge to equilibrium states, in a continuous time course. Sometimes the word "algorithm of neural computation" is used. The actual meaning is a nonlinear dynamical system, instead of that defined in the classic computing theory.

Accordingly, in the neural geometric engine, the information processing is carried out by a special class of nonlinear circuits, the closed loop adaptive circuits. They are the elemental neural processors. An example of the closed loop adaptive circuits in our design are those with real-time adjustable linear combiners, which simulate cortical neurons with dynamical receptive fields. These closed loop adaptive circuits appear similar to Widrow's closed loop adaptive filters. But there is a very fundamental difference. In the adaptive filter concept, the process is defined by the linear operation singled out from the adaptation process. The adaptation process is viewed as a learning process outside the linear filtering. This separation becomes possible because the adaptation process happens in discrete time and the linear filtering process happens in real time. In a nonlinear real (continuous) time adaptive system, it is impossible to separate the linear term from a transient process. Only the equilibrium state is eligible to provide a definite output.

Since feedback signals continuously change the receptive field functions of cortical cells before the closed loop circuit reaches an equilibrium state, the measurement provided by single cells are

transient and not well defined. Having equilibrium states is the property of a nonlinear dynamical system of the neural circuit which cannot be defined by a single neuron. In the neural geometric engine, single neurons are not the elemental processors, although they have well defined functionalities. The closed loop adaptive circuits are the elemental processors.

The digital computer system has a set of logic-arithmetic operations as its computational primitives and builds all processes upon this set of primitive operations. Neural computing also has its functional units. These are the neural computational primitives. Neural computational primitives are the basic building blocks in a closed loop adaptive circuit, each corresponding to certain elemental physical phenomena.

In the neural geometric engine, the summation, multiplication, and exponentiation of analog signals, are chosen to be the computational primitives upon which the closed loop adaptive circuits are built.

In contrast to digital computer systems where memory and processor are separate entities, in brain, memory and processor reside in same network structure. The brain is a hierarchical and distributed system with feedback routes. Different levels of processing and representation of sensor information are able to exhibit increasingly more intrinsic properties of the environment.

In our design, the neural geometric engine is a hierarchical distributed information processing system that includes two levels of functions: the V1 level for extracting affine parameters of local image transform from images, and the V2 level for computing three dimensional motion and surface shape in a viewer-centered coordinate system.

(1) Representation of Visual Information

In the neural geometric engine, visual information is represented as the measurements of image intensity by receptive fields which can be modelled as various spatial derivatives of Gaussian distribution functions or Gabor functions.

Homogeneous intensity does not convey much information about the environment. Visual information is conveyed in spatially oriented contrasts of intensity. A visual field with spatially oriented contrasts of intensity with finite extension can be naturally modelled by directional derivative of Gaussian distributions. As shown in Figure 9, the simple cells in visual cortex are found to have that structure. Receptive field functions of simple cells are the basis functions in cortical representation of visual information, just as bits are the basic form of computer representation of symbolic information.

Another model of simple cells is the Gabor functions (Figure 9). The spatial change of intensity can also be modelled by spatial frequency components. The theory arose from the desire to minimize the joint uncertainty of an event in terms of spatial location and spatial frequency. The

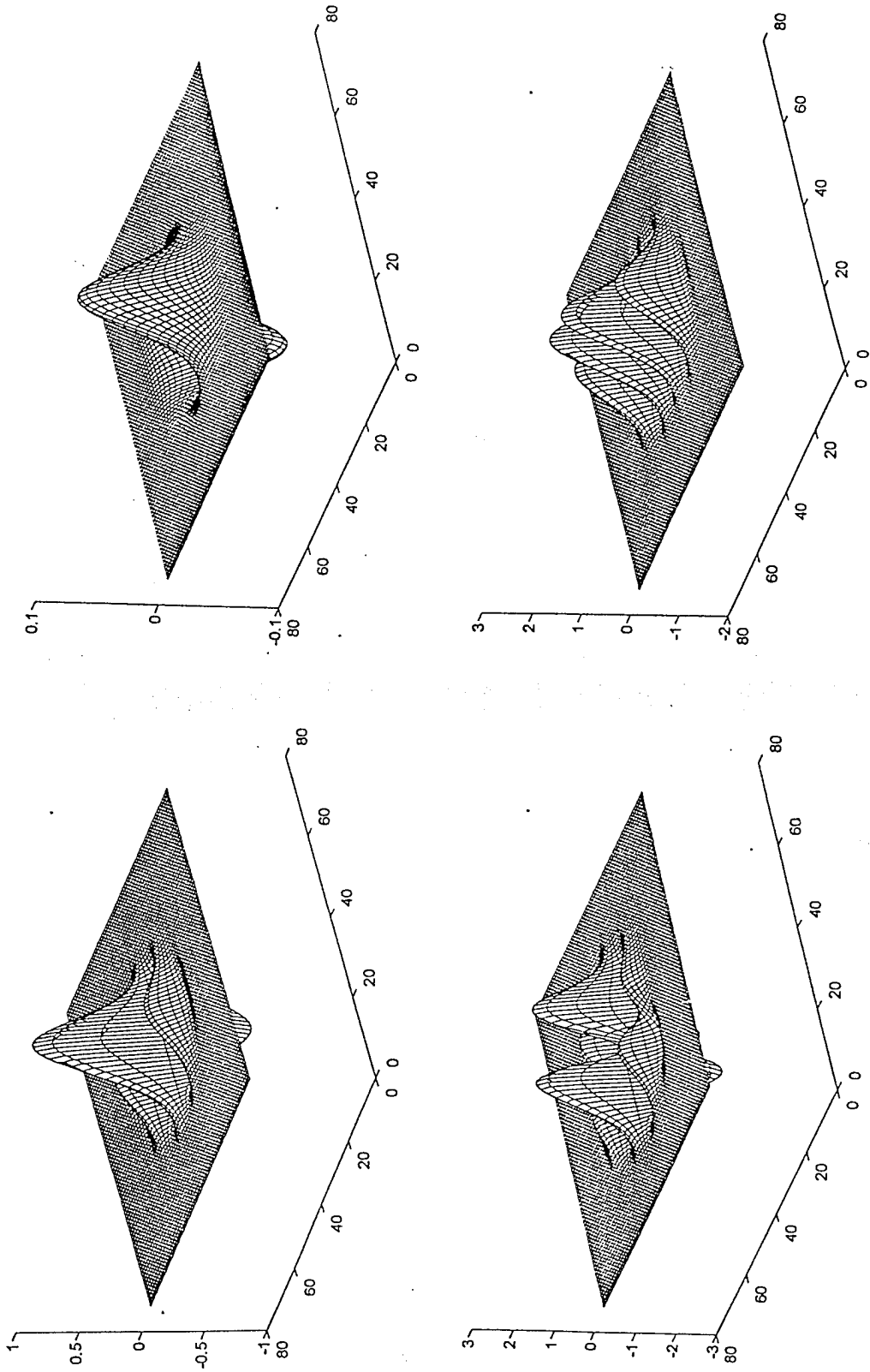


Figure 9. On the top are the perspective views of receptive fields of the simple cell modelled by a Gabor function (left) and the first x-derivative of a Gaussian function (right). The left side of bottom is the rotational Lie germ associated with the Gabor type simple cell, the right side of bottom is the scale Lie germ.

bases functions realizing such a requirement were proven to be the Gabor functions. The Gabor function represents a "quanta of information." Gabor suggested calling the elementary quantum of information a *logon*. A *logon* is a quanta of information in analog signal domain just as a *bit* is a quanta of information in discrete symbol domain.

The Fourier expansion has better convergence properties than the Taylor expansion beyond a narrow neighborhood of a point. The two models, the *directional derivative* and the *Gabor* model, might each have its applications: one for simple parvo cell receptive fields and the other for simple magno cell receptive fields.

There are fundamental differences between the hierarchical representation of visual information in biological systems and the data structures in digital computers. Computer data structures are stored for logical and arithmetic manipulations. In contrast, the biological sensor information processing and representation is a mechanism of adaptation by an animal to its environment. Mead depicted a conceptual arrangement of a single level of neural information processing and representation (Figure 10(a)), which provides some hint of how a neural system organizes visual information in a hierarchical order.

Marr was the first to systematically address the representation issues of visual information. He suggested a modular, hierarchical organization of spatio-geometric information in the visual pathway in three principal representations: (1) the primal sketch, which is concerned with explicit properties of the two dimensional image; (2) the 2 1/2-D sketch, which is a viewer-centered representation of depth and orientation of the visible surfaces and includes contours of discontinuities in these quantities; and (3) the 3-D model representation, whose important feature is that its coordinate system is object-oriented.

Marr's theory clearly depicted the path of information flow from sensor data to invariant object model. The shortcoming of Marr's theory is the lack of an internal dynamical model. The deficiency of Marr's computation theory of vision is particularly obvious in the first level process: detection of zero-crossings. Vision system cannot organize higher level of spatio-geometric description based solely upon the impoverished and isolated zero-crossings without introducing various tricks, strategies, and constraints in processing algorithms to "find feature correspondences" and to infer geometric relation therefrom.

The survival pressure from the environment and the adaptation process has made the primate vision system a geometric engine. The processing of spatio-geometric information must start from the first level of visual cortex. Different from Marr's zero crossing based primal sketch concept, the Lie group model of vision takes the affine Lie transformation group as the "model" which the vision system applies for encoding the spatio-geometric information. That is, the vision system takes affine transform of a local image as a "common" and "acceptable" event, and thus quantitatively measures such a transform.

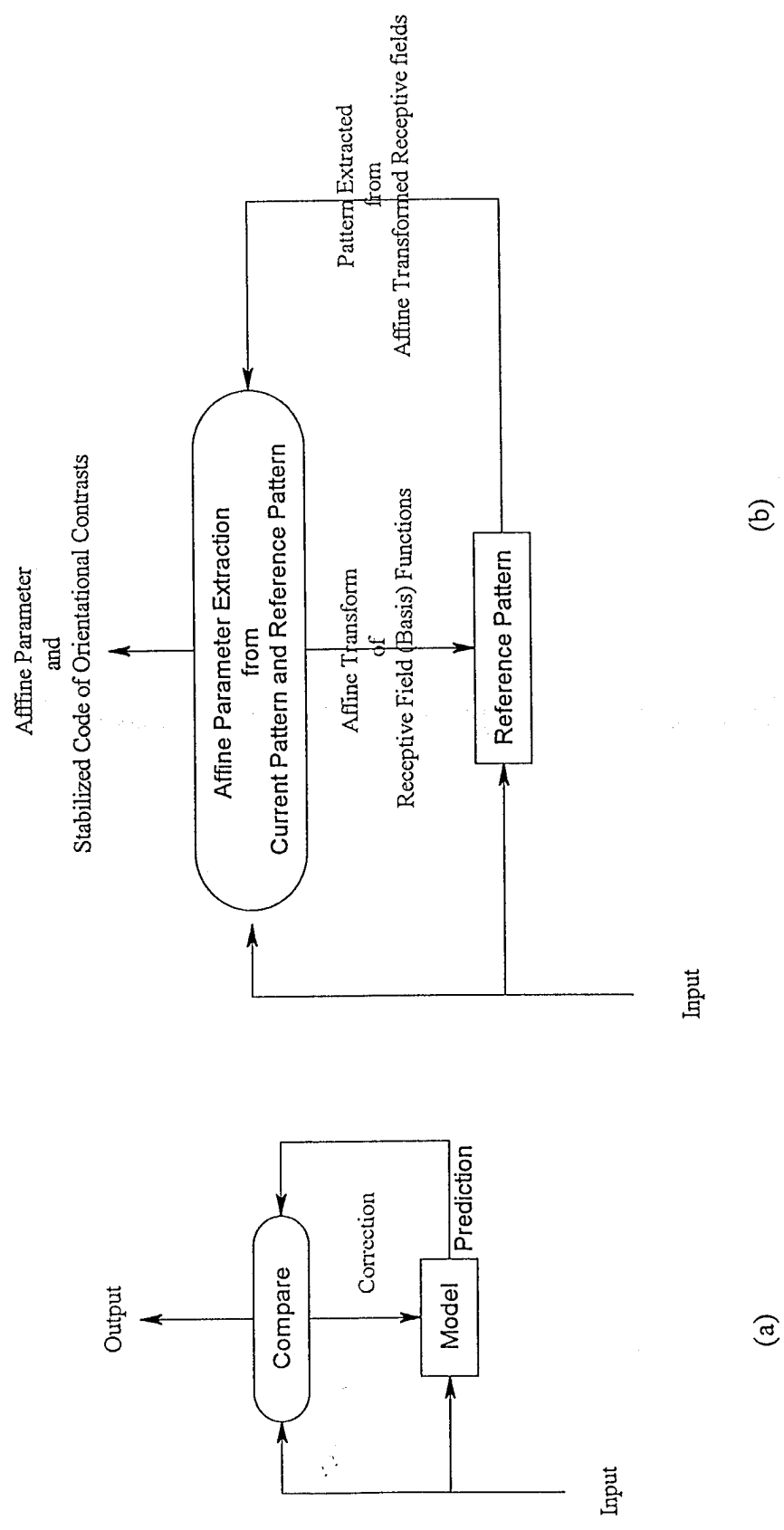


Figure 10. (a) Mead's conceptual arrangement of a single level neural sensor information processing. Sensor information is defined in the context of a process of adaptation, not by the absolute value of signal. (b) The conceptual arrangement of V1 information processing according to Lie group model.

This is conceivable because locally any movement of the eye or object will affine transform images of locally flat surfaces. Affine transforms will be part of a primate's visual experience all the time. For that reason, the V1 will record the affine parameters for the changes, and leave the code of spatially oriented intensity contrast affine invariant (Figure 10(b)). The mechanism of affine Lie group processes is a critical step towards adaptation to a dynamical environment. It makes it possible for an animal to perceive a 3-D geometric world and its motion.

Based on affine parameter measured from first level processing, the viewer-centered surface and motion description is built in the second level representation. Current work on the neural geometric engine will not involve the object-centered 3-D description of the environment.

(2) The Nonlinear Dynamical System for Extracting Affine Parameters

The heart of the neural geometric engine is its elemental "Lie group processor," the particular closed loop adaptive circuit which calculates invariant codes of spatial contrasts and performs measurements of affine parameters of input visual stimulus by setting up the nonlinear dynamical system upon receiving sensor images. The nonlinear dynamical system is the process executed by the neural elemental processor. It represents the most fundamental "algorithm" of our neural system. We will describe it in detail.

Assuming, as much biophysical research has suggested, that the cortical simple cells have Gabor (or directional derivatives of Gaussian) type receptive fields, we will explain how the dynamical receptive field in a closed loop adaptive circuit will facilitate a neural dynamical system that extracts affine parameters upon convergence to equilibrium.

The intensity value of a small image patch $f(x, y)$ of a visible surface is a square integrable (L^2) function: $\iint f^2(x, y) dx dy < \infty$. Here x and y are horizontal and vertical coordinates of pixels. In accordance with the information representation method adopted in the neural geometric engine, the simple cells of different orientation selectivity provides a reference frame for the Hilbert space vector $f(x, y)$. The cortical reference frame (CRF) consists of a set of n , $n \geq 3$, simple cells with receptive field functions $g_i(x, y)$, $i=1, \dots, n$. They are chosen to be rapid descent functions: $g^i \in S$ (for the definition of rapid descent functions, see A.H. Zeemanian "Distribution Theory and Transform Analysis," New York, McGraw-Hill, 1965). They are vectors in the dual space of the L^2 space of the images: Each g_i is a functional on L^2 .

The set of values produced by projecting local intensity image to simple cells in the CRF

$$\gamma^i = \langle g_i, f \rangle, \quad i = 1, \dots, n. \quad (1)$$

provides a CRF representation for the image patch f , where $\langle g_i, f \rangle$ is the Hilbert space inner

product of f and g_i . The linear processors represent the functional g_i , $i = 1, \dots, n$ constitutes a cortical reference frame (CRF).

In equation (1), the n -dimensional vector $(\gamma^1, \dots, \gamma^n)$ is called the cortical coordinate (CC) vector of the local retinal image (briefly, retinal image, or simply image) $f(x, y)$ in the local CRF. Even though the image patch $f(x, y)$ may not be a differential function of the retinal (image plane) coordinates x and y , when the local image $f(x, y)$ undergoes an affine transform:

$$A(\rho) \circ f(x, y) = f(x', y'), \quad (2)$$

where $A(\rho)$ is a 2D affine transform of the image with parameters $\rho = (\rho_1, \dots, \rho_6)$:

$$\begin{aligned} \begin{pmatrix} x' \\ y' \end{pmatrix} &= A(\rho) \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} \rho_1 & \rho_2 \\ \rho_3 & \rho_4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \rho_5 \\ \rho_6 \end{pmatrix}, \end{aligned}$$

the components of the CC-vector are differential functions of the parameter ρ of the 2D affine Lie group:

$$\gamma^i(\rho) = \langle g_i, A(\rho) \circ f \rangle, \quad i = 1, \dots, n.$$

Latter, the Lie derivative of the components of the CC-vector of $f(x, y)$ $\partial \gamma^i(\rho) / \partial \rho_j = \langle g_i, \partial A(\rho) / \partial \rho_j \circ f \rangle$ will be denoted by Ω_j^i .

If instead of using $\rho = (\rho_1, \dots, \rho_6)$ as defined in Equation (2), we use it denote a canonical coordinate of the second kind (see L. Pontrjagin "Topological Groups," Princeton, 1946, Princeton University Press), then Ω_j^i can be calculated as follows:

$$\begin{aligned} \Omega_j^i &= \langle g_i, \partial A(\rho) / \partial \rho_j \circ f \rangle = \langle g_i, X_j A(\rho) \circ f \rangle \\ &= \langle X_j^* \circ g_i, A(\rho(t)) \circ f \rangle \end{aligned}$$

where X_j^* is the Hilbert space conjugate of the infinitesimal generator of the j -th 1-parameter Lie subgroup the 2D affine Lie group $A(2, R)$.

Take as example two image patterns f_d and f_t . The pattern matching in the brain is via γ_d^i and γ_t^i , $i = 1, \dots, n$. And in cortical representation, the *affine invariant distance* between two patterns results from a conjugate (dual) transform $A(\rho)$ on g_i that maximally compensates the affine distance between image data and template:

$$d(A(R, 2); f_d, f_t) = \min_{A(\rho) \in A(R, 2)} \left\{ \sum_{i=1}^n [(\langle A(\rho) \circ g_i, f_d \rangle - \gamma_t^i)^2]^{1/2} \right\} . \quad (3)$$

The spectrum of an image feature are same as that of a template if $f_t \in \text{Traj}(f_d, A(R, 2))$, the affine equivalent class of pattern f_d called the trajectory (or orbit) of f_d under the affine group $A(R, 2)$ defined as:

$$\text{Traj}(f, A(R, 2)) = \{A(\rho) \circ f \mid \rho \in R^6\}.$$

The trajectory is an six dimensional manifold.

This affine invariant distance of patterns and the parameter ρ_0 of the affine transform $A(\rho)$ that maximally compensates the affine distance are calculated via a dynamical process of energy minimization, where the energy function $E(\rho; f_d, f_t)$ is

$$E(\rho; f_d, f_t) = \sum_{i=1}^n (\langle A(\rho) \circ g_i, f_d \rangle - \gamma_t^i)^2. \quad (4)$$

Equipped with analytically calculated Lie derivatives through Lie germs (see Figure 9), it is straightforward to construct a dynamical process to determine an affine invariant representation of data relative to a template by minimizing the energy function. A gradient system or a Newton-Raphson system are candidates for such dynamical systems. For numerical execution of the dynamical system, the Newton-Raphson scheme converges rapidly when the solution is in a neighborhood of an initial guess.

In our design, the closed loop adaptive circuit containing simple cells and Lie germs and the feedback control is called a **Lie group processor** (see Figure 11). The design of the Lie group processor simulates the hypercolumn structure in visual cortex which contains many orientation specific microcolumns. The Lie group processor contains n different orientation specific units (see Figure 12). The simple cells in these specific orientation units constitutes a cortical reference frame (CRF) for coding the local image intensity distributions. The intrinsic neurons are responsible for affine transforming the CRF to keep the CC-code stable, and the binocular CC-code be fused.

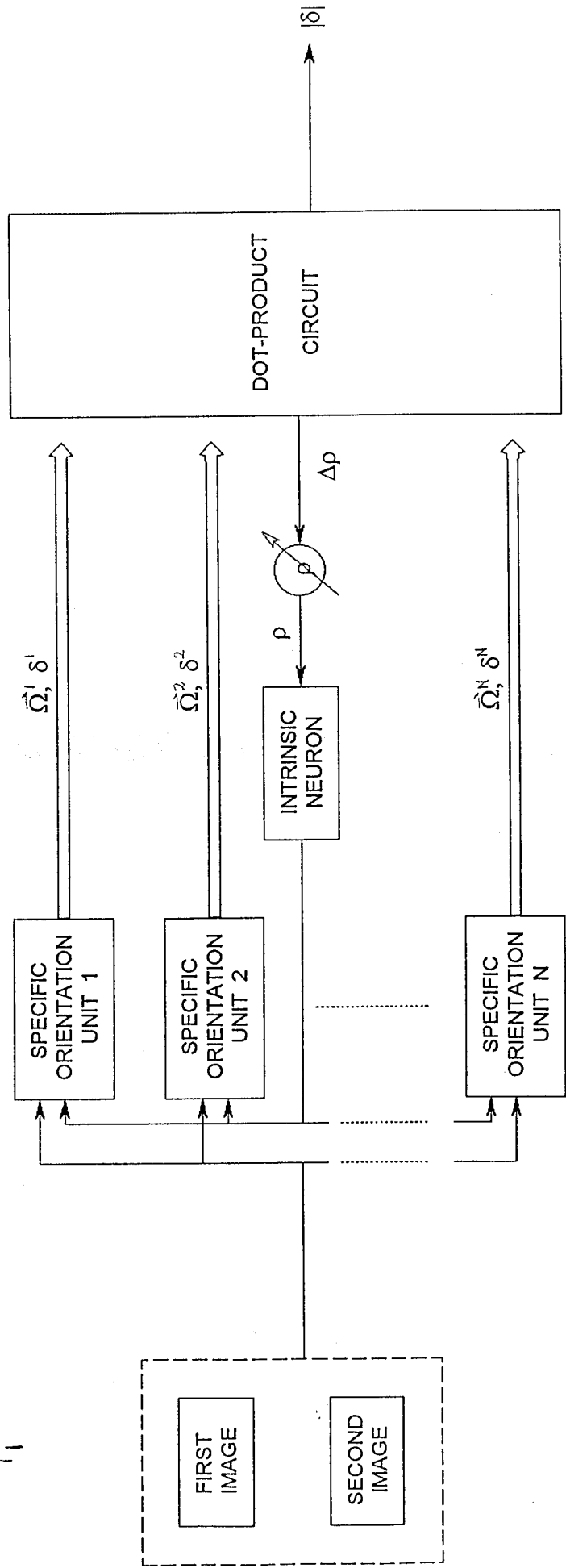


Figure 11. The Lie Group Processor, a closed Loop adaptive circuit for determining Affine parameters. The circuit returns affine parameters when the dynamical system reaches the minimum energy state.

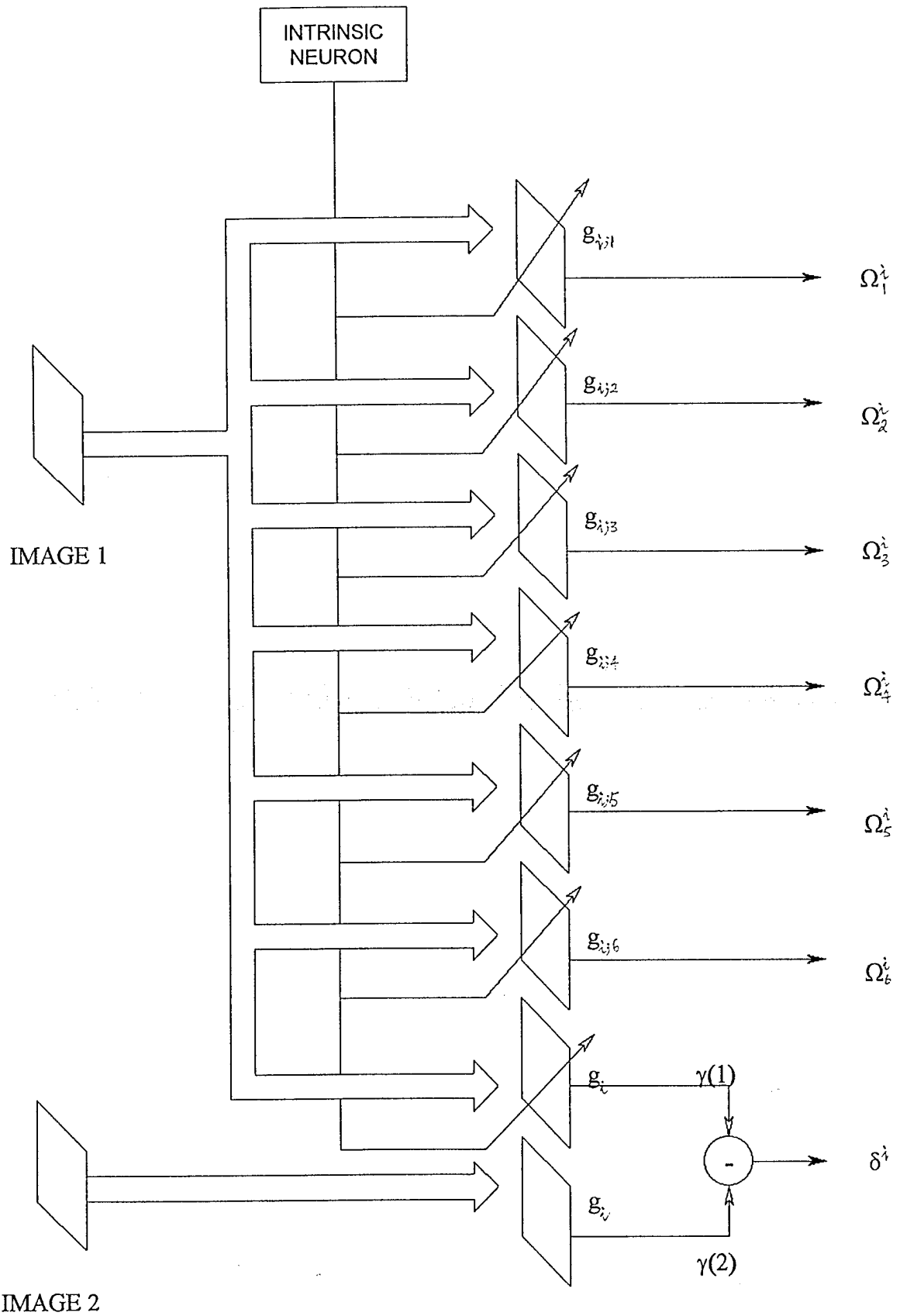


Figure 12. Linear process and receptive field transform in a specific orientation unit i .

The affine transform operation of the intrinsic neurons is controlled by feedback signal from the remaining differences of (binocular) CC-codes δ^i and the Lie derivatives Ω_j^i . The extraction of affine transform parameter at the minimum energy state is accompanied with the computation of (binocular or motion) invariant CC-code of the intensity pattern.

In our Lie group model, Lie group processors are the basic circuits in V1, and the representation of visual information in V1 has two parts: affine invariant *CC-vector* and *affine parameters*. This is coincident to Gibson's view that the vision system picks up two kinds of information: *optic array* and its *transformation*. It is very different from Marr's primal sketch, and all those based on a feature detection paradigm, which only cares for the static contrasts of intensity.

As matter of fact, all the later geometric information processing is from the *transformation* part. For example, the parameter of shift in binocular fusion affine transform determines the range from the viewer. In Figure 13, using a simplified Lie group model, local shift parameters are computed from two consecutive photos (shown on the top) taken from an airplane. The result show at the bottom indicates the range at each point. The shift parameter together with other two parameters further determines the surface 3-D orientation, etc. In this sense, extraction of the affine parameter is the starting point of spatio-geometric information processing.

(3) Neural Computational Primitives of Lie Group Processors

In a computer program, not just any sector of a code defines a process. The criterion for being an individual process is if it defines an input-output relation. In neural processing, the operational meaning will be: does the network define a dynamical system that leads to some equilibrium state? The sensor signal input to a neural system generates a disturbance of the system and initiates a dynamical process which may lead to some equilibrium state. If the dynamical system leads to a stable equilibrium state, it defines a process.

According to the Lie group model, A V1 level process is not defined by linear processes performed by cell receptive fields. (This is different from the "feature detector" doctrine, in which V1 processes are defined by the linear "orientation selective" cells and other selective response cells). A V1 process is a dynamical process participated by *affine Lie group elements* (intrinsic neurons) which help fuse the binocular image and compensate motion affine effect by transforming the receptive fields of the linear "orientation selective" cells. That is, the intrinsic neurons are functioning as agents for the cortical reference frame transformation. During the process, the receptive fields and their output signals are transient, until they reach a minimum energy state.

The computational primitives are the "elemental forces" which participate in the dynamical process, collectively generating and changing the transient phase vector in a nonlinear dynamical system. The neural representation and processing of visual information is determined by the structure and real-time dynamics of the receptive fields of cortical relay neurons, as well as the

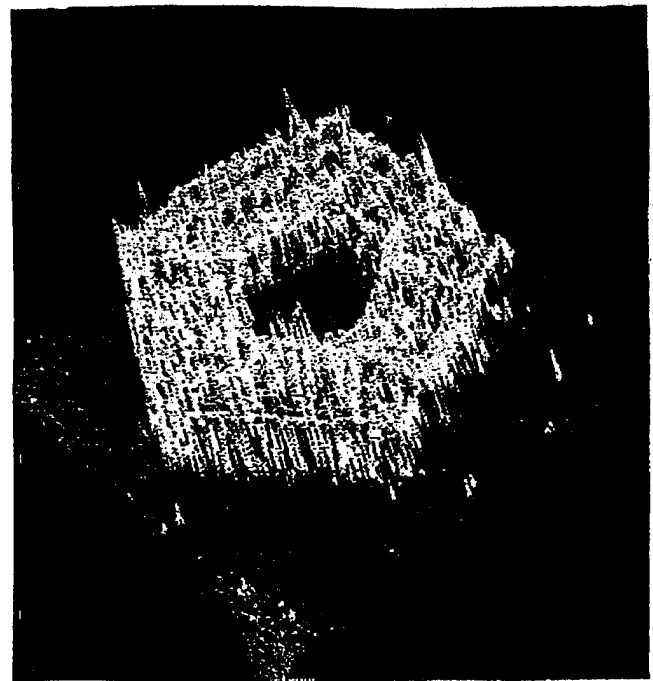
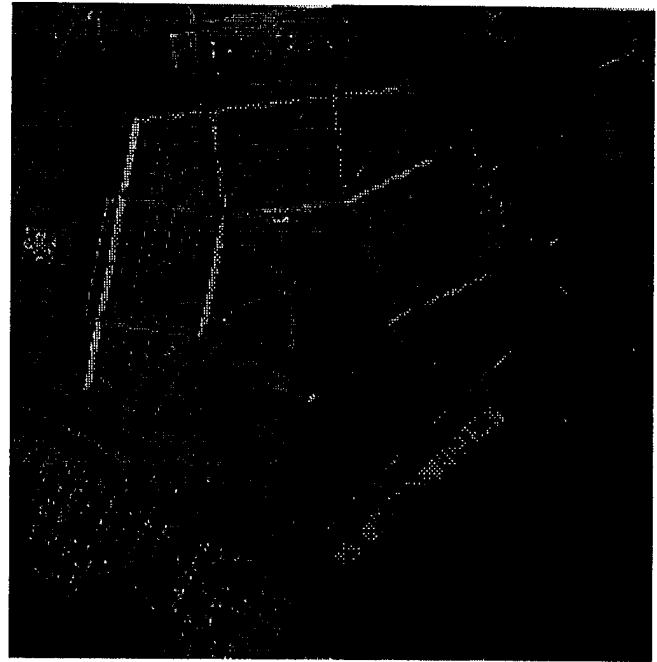
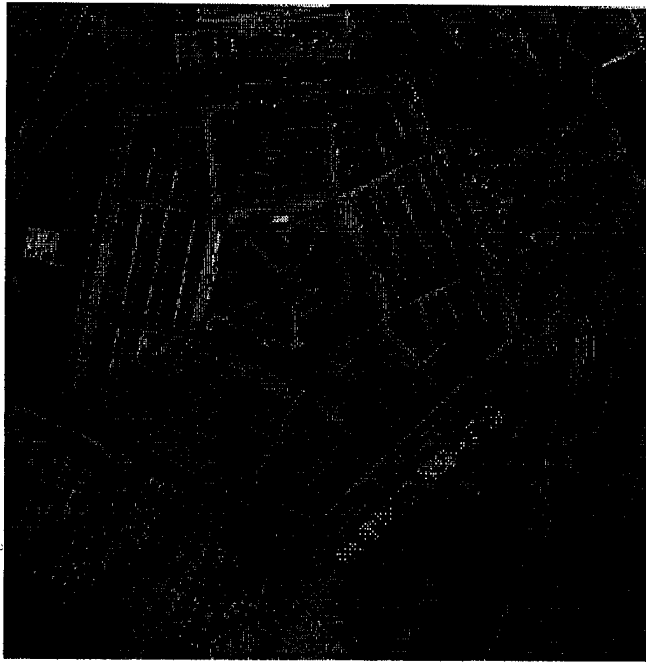


Figure 13. On the top is a pair of Pentagon images taken from above. The left side of bottom is the map of shifts between two images, generated by the Lie group model neural system employing only shift-parameter Lie germs, in the form of intensity image. The right side is the three dimensional display of the shifts, which is proportional to the ranges from the camera.

interaction and participation of intrinsic neurons.

Separated from the feedback network, the operations of simple cells can be viewed as linear. The simple cells act as linear combiners. However, as part of nonlinear dynamical process, the receptive fields of the cells are transformed in real-time. Thus, the overall nonlinear dynamical process involves more than the primitive operations of a linear process, namely multiplication and summation. The process also involves as its primitive function the *exponential mapping* for transforming the receptive fields, since the receptive fields take the *Gaussian distribution function* as the basic form of spatial extension.

There are profound reasons for the Gaussian distribution function being taken as nature's choice for the basic form of spatial extension of receptive fields. (For example, the requirement of minimum joint uncertainty of spatial location and spatial frequency leads to the form of Gabor functions which are Gaussian modulated harmonic functions.) In neural processing of spatio-temporal information, various types of receptive fields have forms derived from this basic Gaussian distribution form of spatial extension. The implication of this particular form to the neural geometric engine architecture is the inclusion of the *exponential function* in the primitive operation set along with *multiplication* and *summation*.

(4) The Organization of The Neural Geometric Engine

The current design of the engine has two levels of processing: (1) extract affine parameters of local transforms from images, and (2) compute three dimensional motion and shape in a viewer-centered coordinate system. These two levels of processing correspond to the magno stream processing of areas V1 and V2 in primate visual cortex. The function of V1 processing is to extract sensory parameters from images, and the function of V2 processing is to further infer 3-D geometric and kinetic parameters of the visible surface from the sensory parameters. Both levels of the early vision process are local and driven by sensory data.

The primate's vision system has been highly developed for accurate perception of three dimensional shape and object motion. The perception of 3-D motion and shape of objects do not just emerge in some high level specialized visual process areas. Rather, it is supported by expanded lower level sensory data processing.

According to our Lie group model, V1 processing in the primate's visual system is significantly different from the visual processing in lower forms, such as the frog's moving feature detection. Frog's vision system sees no difference between a far away big object and a nearby small bug and gives same response. Primate's vision system sees same object in a close distance or in a far distance. This gives primate extra flexibility to respond to its environment.

The tremendous bottom structure of V1 does not exist only for performing simple tasks by the "selective response" cells such as feature detectors or motion detectors. The structure is a large

collection of closed loop adaptive circuit modules supporting nonlinear processes (i.e. analog neural computations). Composed of many types of linear cells and intrinsic cells, these Lie group modules are able to perform sophisticated measurements of affine parameters involved in binocular and motion vision, while maintaining stable representation to same object.

Without the broader foundation of the lower level processes, the higher level processes would be baseless. As a matter of fact, area V1 of visual cortex is the largest of all the cortical areas of the macaque's brain (15% of all neocortex). The receptive fields of typical V1 cells receive input signals from 800 to several thousand retinal ganglions for local processing of visual information.

In some sense, the evolution of the primate's vision system not only created advanced high level visual areas, but more importantly, created a much more sophisticated lower level visual area. In order to be able to maintain a stable response to same object in motion, the primate's vision system has a large facility to support the transformable local cortical reference frames, *i.e.*, to make the receptive fields of linear cells in hypercolumns dynamical. This extra structure also facilitates the parameter measurements of the affine transforms. In contrast, frog's vision system only has a rigid reference frame.

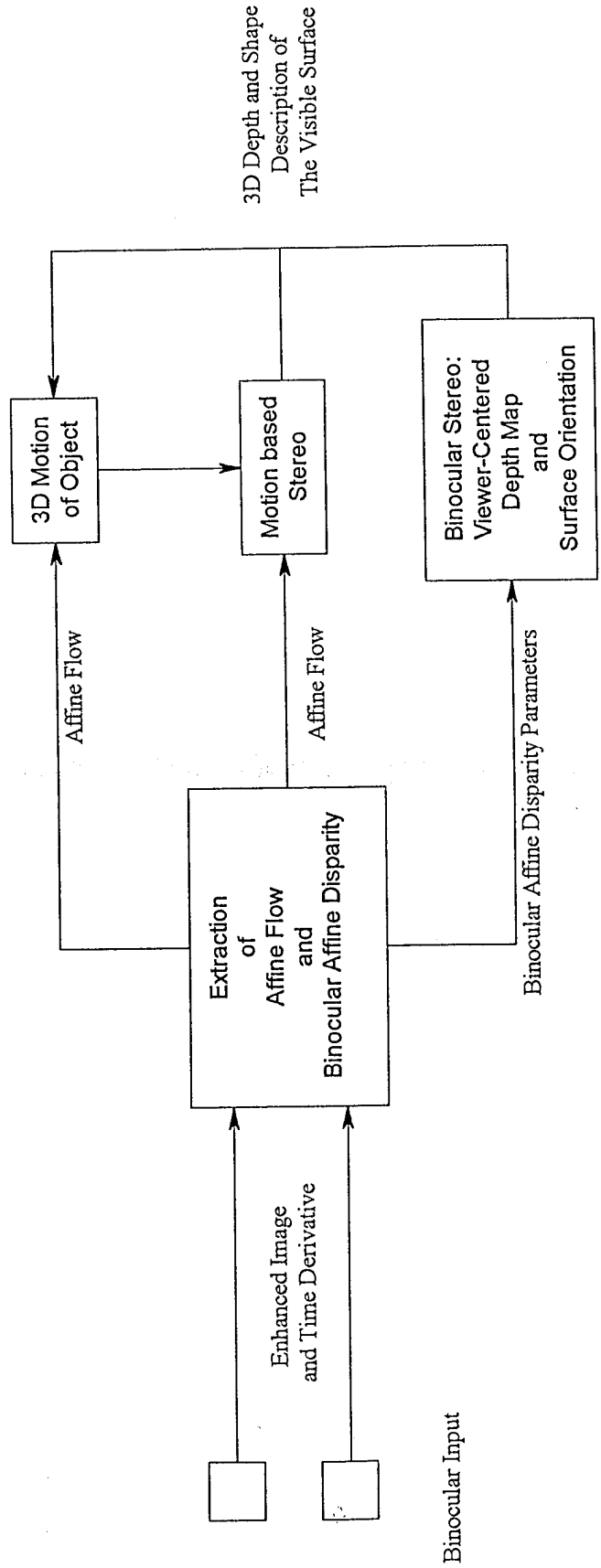
Most vision theories are based upon the concept of feature detectors. The prototype of the feature detector concept is the classic concept of static receptive fields formed in 60s, such as described in Hubel and Wiesel's work. It was only after 80s that the dynamical properties of cortical receptive fields become center of attention of neurobiological research. The vision system with rigid receptive fields, such as frog's, has very little capability to represent spatial information, mainly limited to retinotopic positions of features. It is sufficient for a frog to live in its limited environment. But for artificial vision system designers, this has caused serious problem.

Because no significant spatio-geometric information is represented in the zero-crossings or other features, some outside viewer (or ad hoc heuristic computer program) must supply it by *finding the feature correspondences*. The difference between this proposed geometric engine and other machine vision systems and "image understanding systems" is mainly in the lower processing level. It is the unique internal dynamics embedded in the lower processing level that makes the neural geometric engine an autonomous visual engine.

Thus, different from other computer based vision systems, the neural geometric engine contains no feature detection. It has two levels of computation after the sensor input level (see Figure 14): the affine transform analysis level, and the viewer-centered three dimensional modelling level.

3. The Design of Digital Version Neural Geometric Engine

Digital implementation of a neural geometric engine means using digital computing system to



Sensor Input Level Level 1 Process: Sensory Parameter Extraction Level 2 Process: Viewer Centered 3D Shape and Motion

Figure 14. The neural geometric engine includes two cortical processing levels: Level 1 is for sensory parameter extraction, based on the affine Lie group model; Level 2 is the 3D model build up, including motion and shape.

perform numeric simulation of the cortical process of vision, in contrast to the analog implementation which directly mimic the cortical process. However, certain parallel distributed processing characteristics still can be retained.

A pure software computer simulation can use workstations. The workstations such as Sun SPARC, HP J210 series, and ALPHA offer 50 to 200 MIPS processing. However, simply using a high performance workstation as the computing engine has several drawbacks:

- I. Complex operating systems can occupy as much as 60% of CPU's processing time.
- II. Sophisticated graphics display and graphical user interface demand extensive CPU processing.
- III. Real-time applications require an additional software layer, which must coordinate the disk, graphics, and peripheral I/O, host-to-data acquisition, and data-to-host transfers.

Those drawbacks can be easily bypassed by integrating the workstation with dedicated parallel processing host-based hardware. Such hardware, DSP, can accelerate cycle time for CPU (central processing unit).

During the past decade, DSP performance increased from 5 MIPS (million instructions per second) in early 80s to over 2 BIPS (billion instructions per second) today. The development of DSP chip technology has made possible to implementing real-time or near real-time processing for the neural geometric engine. Particularly, the advanced DSP chips commercially developed by companies of Motorola, Analog Device, AT&T, and TI are suitable for the tasks.

In this implementation, the specialized DSP will be 100% dedicated to the neural geometric engine computation. Multiple DSP chips working in parallel will provide several hundreds to several thousands times of computing power of a high end workstation, and will make possible for real-time processing of the neural geometric engine.

Figure 15(a) shows the block diagram architecture for a C40 DSP chip which includes 32-bit floating -point parallel central processing unit with some multichannel direct-memory-access (DMA) co-processor, six communication ports, memory, program cache, 32-bit global and local memory buses, two times, and an analysis module. This architecture is especially suitable for parallel multi processing system, which meets these criteria:

- I. High processing speed;
- II. A large number of high-speed DMA channels supported links;
- III. Ease in load balancing (even processing distribution over all the processors);

IV. Easily configurative and incremental expandable architecture;

V. Ease in programming via multitasking kernels and multi processing program support;

VI. High speed I/O.

Figure 15(b) shows a building block of a digital implementation of the neural geometric engine using parallel DSP computing. The structure of Figure 15(b) is widely used in parallel processing, where large data is segmented and decomposed. There are five DSP nodes in the building block. The DSP node on the top is called level 1, which can have only four level 2 nodes, because the communication ports are limited. This DSP structure is designed to implement a "specific orientation unit" of the neural geometric engine as shown in Figures 11 and 12. There are six specific orientation units and one DOT-product circuit in the neural geometric engine. Six nodes of level one and one node of level 0 consists of thirty-one C40 DSP chips to implement one neural geometric engine shown in Figure 16.

Two advantages will benefit the DSP implementation of the neural geometric engine: DSP offers DMA and CPU operations over the link to reduce communication overhead for large amount of data. Also DSP independently processes the data without slowing down the others. Thus the parallel processing and high data throughput make the DSP a suitable digital means.

4. Issues of Analog VLSI Implementation of Neural Geometric Engine

The neural geometric engine architecture as above described can be *most naturally* implemented using analog VLSI technology. All three computational primitives correspond to fundamental physical phenomena in silicon circuits; Analog signals from sensor can be sent to artificial linear cell's "receptive fields" to be processed and represented in the system without having go through all the binary coding and processing. An energy minimization process will occur in a continuous time physical process in the circuit, without suffering the convergence problem caused by discrete time and numeric round-off error.

As described before, the architecture of neural geometric engine is to mimic primate's visual cortex as we understood via Lie group model. The single most important and central concept for understanding the functions of visual cortex is the **receptive field**. The relay cells use their receptive fields process visual information and generate a cortical representation of it. The intrinsic neuron uses their arborized axon to transform relay cell's receptive fields, etc. The receptive field structure makes neurons able to process visual information. Receptive fields are the elements and "bits" of the neural geometric engine. They are the building bricks. A prominent and universal feature carried in by the receptive field structure of neurons is the **extensive, two dimensional, mathematically defined integration** of signals in each basic processing step. This is very different from most of the artificial neural networks, as well as artificial retinas. This is an essential feature of the primate's primal **visual cortex**.

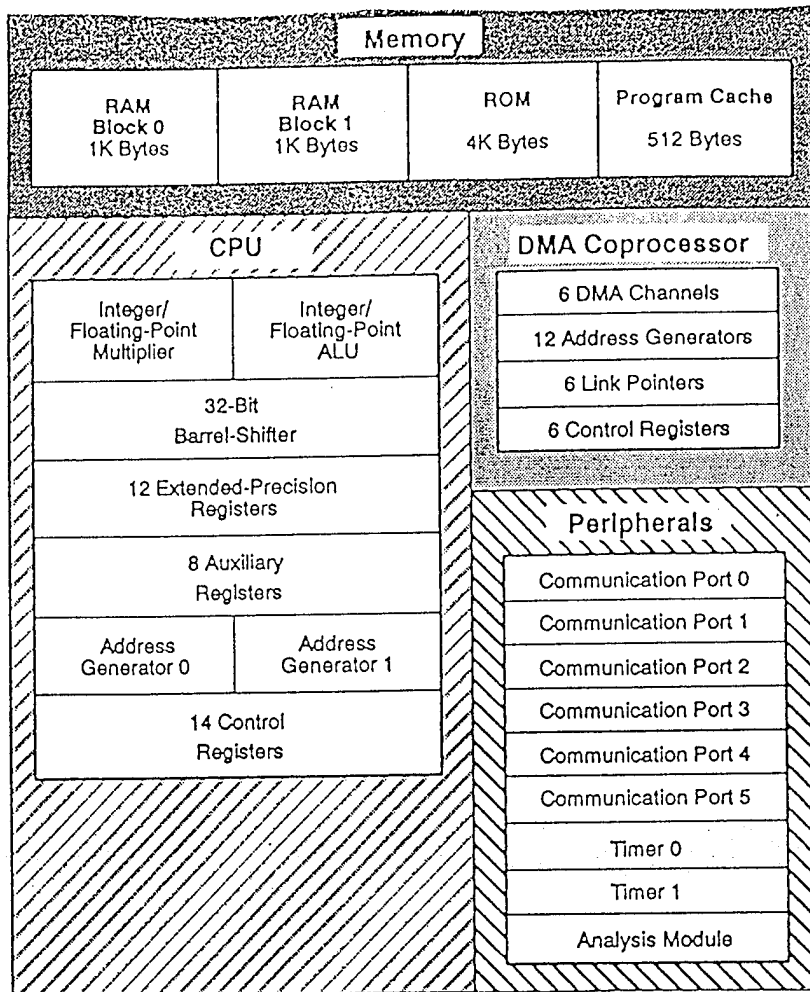


Figure 15 (a). DSP Block Diagram

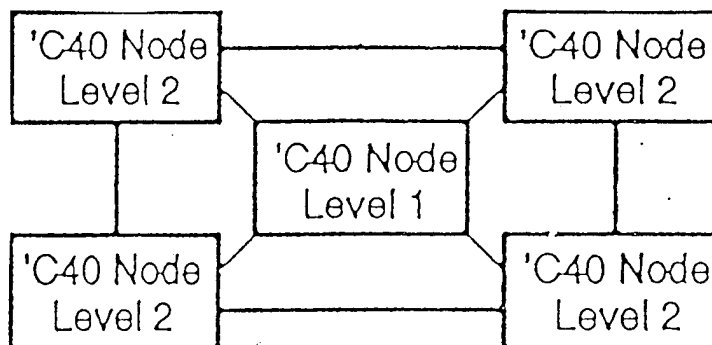


Figure 15 (b). Parallel Structure Block With Five DSP Nodes

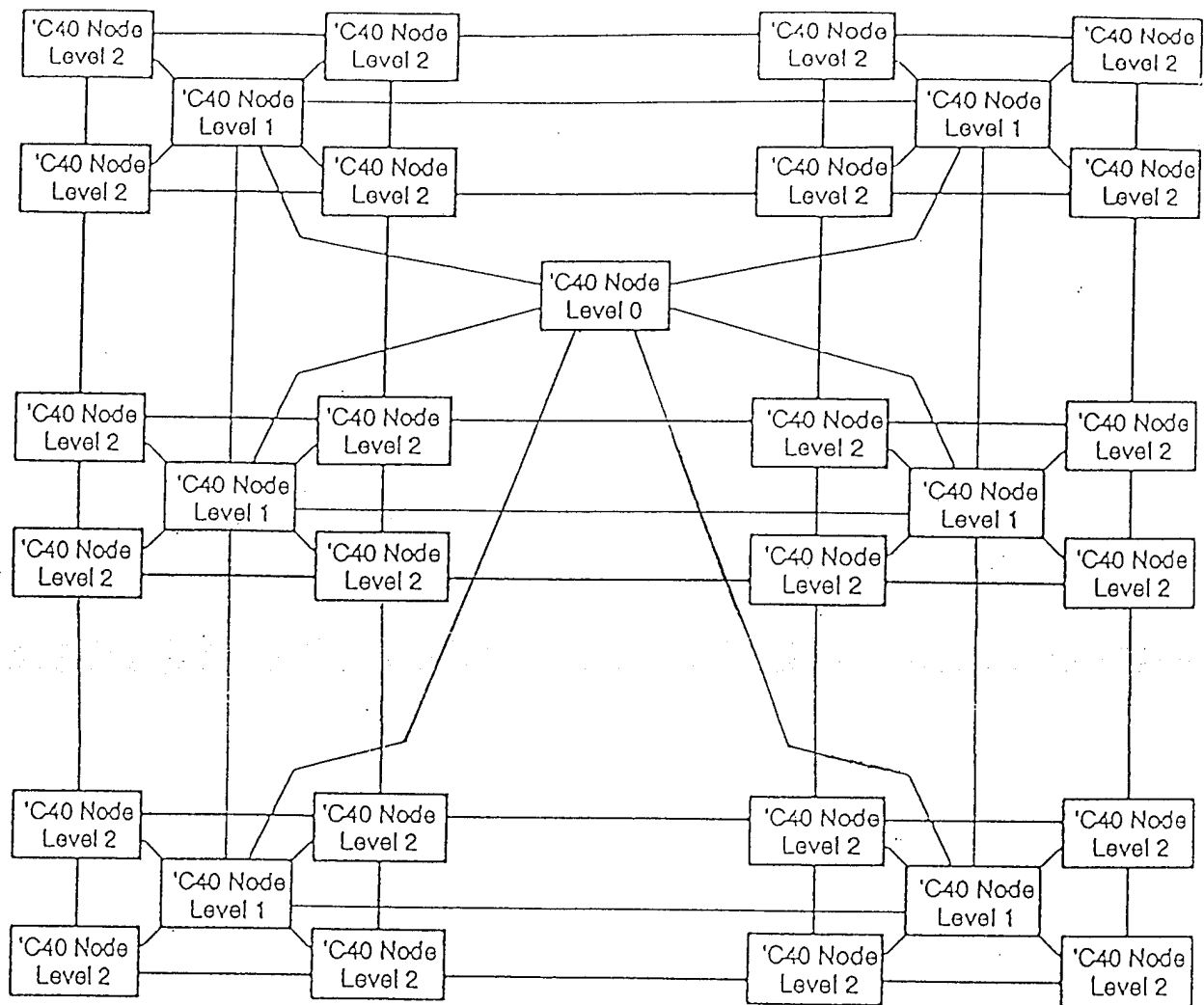


Figure 16. The Impiementation Of Nural Geometric Engine With DSP Parallel Structure

The technology described by Mead's group and many other groups working on **artificial retinas** is basically a two dimensional neural structure. However, the brain, as correctly pointed out by Mead, is a $2 + \epsilon$ structure. Simple cell is to represent the spatially oriented contrasts. They represents *logon's* of the two dimensional sensor images, and not pointwise pixels. For that purpose, nodes in a quite extensive area will be included in a linear combination operation. The cortical receptive fields are highly overlapped. A strict two dimensional structure will face a difficult wire crossing problem. A third dimension is necessary as the processing arising to higher levels, because each level requires integration of a substantial number of nodes from the level below, not just from some immediate neighbors. The first two dimensions of the structure are necessary for representing the extension and resolution of images. The third (ϵ) dimension is necessary for the levels of processing. Without the third dimension, the natural parallelism in visual processing will be eliminated substantially and the flow of visual information will be bottlenecked.

The difference between the technology suitable for an artificial retina and the technology suitable for an artificial cortex is the third (ϵ) dimension. To alleviate this computational bottleneck Irvine Sensors has developed a three dimensional artificial neural network, 3DANN. Stacks of two dimensional structures are highly interconnected. Irvine Sensor's 3DANN has the computing power to compare 260 million templates to an incoming image every second with a power dissipation of less than 2W. Despite the substantial differences between the architecture and functions of the neural geometric engine and that of the 3DANN neural network, the success of 3DANN indicates that all the necessary components of technology for implementing the neural geometric engine, an artificial visual cortex capable of spatio-geometric perception, are already available, or within the reach.

The analog VLSI technology provides a viable approach to creating a computing system that distinguishes itself from the existing supercomputers by many orders of magnitudes in terms of computing power, physical size reduction, energy efficiency, and robustness.

5. Three Major Fields of Applications

The neural geometric engine is not only a new way of providing the computing horse power. It is not only a new way of computing. Most importantly, *it computes information that has never before been computable by machines*: The affine invariant CC-vector of image intensity and the parameters of affine transforms between image parts. The great query of Pitts and McCulloch, "How we know universals" was not answered by AI research and neural network research. The significance of computing this type of visual information is that without it our efforts at object recognition are baseless: In order to *recognize* something we have to *perceive* it properly. The failure of the image understanding approach is rooted in its methodology of trying bypass the perceptual process, not just the lack of enough computing power, although it is true that to get this critical piece of information is very costly in terms of digital computing.

The Neural Geometric Engine will change the way of thinking that currently dominates the design and development of algorithms and computer systems for stereo vision, pattern recognition, and sensor fusion applications. Innovations in these application fields will be developed as the results of applications of the neural geometric engine.

1. Stereo Vision

Surface shape can be derived from binocular stereo images or successive images taken from a sensor system on a moving platform. Current state of the art only calculates the shift disparity and range map of the visible surface through a feature matching process. The local affine parameters extracted from binocular stereo images also determine the orientations of surfaces at each visual direction. This gives a complete description of the shape of the surface. The surface orientation information will be useful for various military and industrial applications.

2. Fusion of Multiple Images

The second field of applications is fusion of multiple images. Typical examples are binocular fusion and image registering. Usually, the differences between images subject to fusion cannot be removed by simple shift operations. It involves local affine changes, such as scale, rotation, and shear transforms. Local geometric correction is needed in order to register multiple images with geometric deformation, or mosaic images taken from different positions into a large view of a scene.

Conventional local geometric compensation processes takes substantial time because they use brute force "rubber sheeting". In many real life applications, image fusion must be performed in real-time. The supply of local affine parameters by the neural geometric engine will advance the state of the art of this application field.

3. ATR and Automated Screening of Image Data

The third field of applications is automatic target recognition and automated screening of large numbers of image data. In these applications, computer systems are employed to detect, classify, and recognize image features of targets of interest. In real life applications, the sensor image data is always subject to variations of scale, rotation, translation, and shear. While feature matching and classification are quite straightforward processes, the geometric variance in data poses great difficulties for ATR and target feature detection in image screening. Image geometric variances may cause detection miss, classification miss, or lead to false dismiss. Usually, ATR with image geometric variances requires tremendous computing resource and computing time. Naturally, a breakthrough in handling the geometric variances will greatly advance the state of the art of ATR and automated image screening technologies.

Scientific & Technical Reports Summary - Thomas Tsao - DASG-95-C-0080

The novel method of automatic target recognition applies a geometric compensation processor prior to image feature matching (see Figure 17). The function of the geometric compensation processor is to remove the image feature variances, such as scale, rotation, shear, and translation changes, and reduce the image feature to a "standard presentation" before matching the templates for the purpose of detection and classification. Neural geometric engine is expected to be applied to substantially reduce the number of templates and matching, and reduce the error rates.

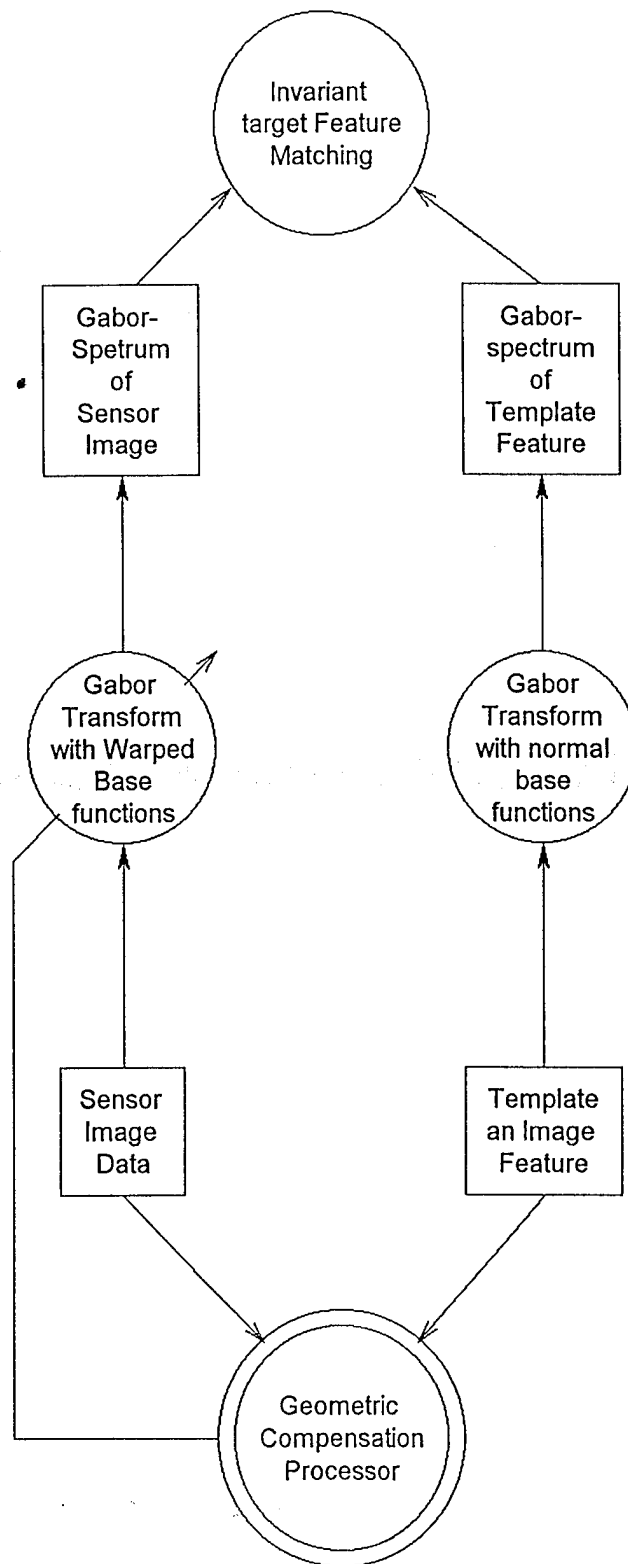


Figure 17. Apply Geometric Compensation Processor to the Gabor base functions will allow invariant target feature matching when data variances including changes of scale, rotation, shear, and translation.

5. Conclusions

Giving Gibson's smart sensor concept of spatial vision a computational theory and implementation, neural geometric engine is the first artificial vision system in which the basic process is analytically formulated. Compared with the feature matching based computer vision approaches, the analytical method has compelling advantage in reducing computational complexity and uncertainty, achieving high accuracy and robustness.

Various new computer architectures have achieved impressive progress in speed and storage. They provided new capacity to image and signal processing. The neural geometric engine is different from these computers in basic information representation method, processor concept, computational primitives, and organization. It is a neural computing system and can be implemented in analog VLSI to reach the level of speed, compactness, and energy savings of the analog computing. Moreover, as a neural computing system, it not provides the computing power, but also provides the effective "algorithms" for the early vision process, without which a powerful computer is only a helpless giant.

Even in digital implementation, the neural "algorithm" for vision process is different from a computer vision algorithm. It is a digital simulation of the deterministic analog process in neural circuit, while computer vision algorithm, with feature matching as the central piece, is an common sense method of image data processing. The common sense method, supported by various ad hoc strategies (or "knowledge"), are usually very fragile.

The neural geometric engine is different from most neural networks. It is not a piece of associative memory. It simulates the spatio-geometric information processing neural circuits in primate's visual cortex. Compared with neural networks, the neural geometric engine carries image analysis functions in different ways and different aspects. The spatio-geometric information extracted from neural geometric engine can be used for various passive sensor based measurements and modelling. It also opens a new way of invariant object recognition.

6. Implications for Further Researches

The most important implication of the neural geometric engine research is that it will change the way of thinking that currently dominates the design and development of algorithms and computer systems for stereo vision, pattern recognition, and sensor fusion applications. It implies a revolution in artificial vision research.

The currently dominant **information processing** paradigm in vision research is based on a deep belief that the spatio-geometric relation must be derived from some measurement from images, and the procedure of executing such measurement is **feature detection** followed by **feature matching**. The concept of **matching** become so predominant after three decades practicing in vision, although without much success, that few were questioning on it. To many who are working on the field, the only thing can be done is to make the feature detection and feature matching procedure more effective, and faster.

We have shown that in biological vision system, spatio-geometric relation is measured in a real-time process of dynamical warping and shifting of receptive fields for maintaining the stable representation of moving object or fusion binocular images. The real-time measurement of spatio-geometric relation is **not happening in image domain**. It happens **in the dual space of images**, the space of reference vectors that the brain provided as basis for representing image data. The geometric measurement is in the dual space via a process of "adapting" to motion or binocular disparity. The measurement is accurate and robust. The process of measurement is determinate and can be described by a dynamical system using Lie derivatives, a Lie group model.

The neural geometric engine will be the first of its kind in artificial vision systems, as well as in artificial neural systems. The implementation of the neural geometric engine will make possible the research and development of innovative methods of ATR and automatic image screening, stereo vision, and image fusion, which all depend on geometric computation from sensor images. It is anticipated that the actual use of the neural geometric engine through these three application fields will stimulate more interesting research topics and lead to more development of artificial vision systems and artificial neural systems.

7. Special Comments

1. Marr's theory is best represented in his book "Vision", the bible of computational vision research. Gibson's theory is best represented in his book "The Ecological Approach to Visual Perception".
2. The Lie group model of early vision and the neural geometric engine is not an improvement of current art. It is not even an innovation of information processing method for image understanding.
3. The Lie group model of early vision and the neural geometric engine changes the very basic concept underlies all the algorithm design and system concept in image understanding, the so-called information processing method.
4. The Lie group model is not a description of a method of geometric computing from the images, but a description of the dynamics of the smart sensor itself, a description of the process for adaptively maintaining invariant representations of objects in the brain.
5. In neural geometric engine, the "algorithm" (computational structure) determines the architecture of the computing system, and the computing system implements the algorithms. They are two faces of a coin. This is very different from that of digital algorithm design, which relatively independent of the architecture design.