Lie Group Model Neuromorphic Geometric Engine for Real-time Terrain Reconstruction from Stereoscopic Aerial Photos

Tien-Ren Tsao[a] and Doris Tsao[b]

[a]CompuSensor Technology Corporation, Silver Spring, Maryland
[b]Neuroscience Department, Harvard Medical School, Boston, Massachusetts

ABSTRACT

In the 1980's, neurobiologist suggested a simple mechanism in primate visual cortex for maintaining a stable and invariant representation of a moving object: The receptive field of visual neurons has real-time transforms in response to motion, to maintain a stable representation. When the visual stimulus is changed due to motion, the geometric transform of the stimulus triggers a dual transform of the receptive field. This dual transform in the receptive fields compensates geometric variation in the stimulus. This process can be modelled using a Lie group method. The massive array of affine parameter sensing circuits will function as a smart sensor tightly coupled to the passive imaging sensor (retina). Neural geometric engine is a neuromorphic computing device simulating our Lie group model of spatial perception of primate's primal visual cortex. We have developed the computer simulation and experimented on realistic and synthetic image data, and performed a preliminary research of using analog VLSI technology for implementation of the neural geometric engine. We have benchmark tested on DMA's terrain data with their result and have built an analog integrated circuit to verify the computational structure of the engine. When fully implemented on ANALOG VLSI chip, we will be able to accurately reconstruct 3-D terrain surface in real-time from stereoscopic imagery.

Keywords: Lie group Model of Early vision, Stereoscopic imagery, 3-D model of Terrain, Real-time reconstruction, Analog Neuromorphic device

1 INTRODUCTION

In 1995 SPIE Orlando conference, a Lie group model of neural computing for motion and surface recovery was presented1,2. This paper presents the progress and breakthroughs we have made in developing a neural computing device for 3-D surface perception. There are major theoretical differences of our approach and the mainstream computer vision approaches to 3-D recovery of surface from stereoscopic imagery. We will point out the advantages of our new approach in terms accuracy and computational complexity, and how our approach leads to a neuromorphic implementation.

2 WHAT IS NEW OF OUR APPROACH COMPARED WITH OTHERS? WHAT IS THE ADVANTAGES?

It is well known that 3-D surface information is mainly made available by binocular disparity or motion parallax. It is also known that local surface images in binocular or motion stereo are in general skewed from place to place. The process of fusing stereoscopic local images must simultaneously compensate shift and skew between stereoscopic images, from place to place. It is important to notice that except for a frontal surface patch, a shift operation alone will not fuse a stereoscopic image pair. There are some mathematical misconceptions which assume that by simply keeping the sizes of image patches small, a simple shift operation will be adequate for image matching.

For example, if an image patch is half the scale of its stereoscopically correspondent patch, then any part will be half the scale of the correspondent part. A shift operation alone will never match the two patterns, no matter how small the image patch is. The shift disparity model in stereo vision and the optical flow model in motion vision are simple but inadequate. We know that a logic system with only one operation, say the AND logic operation, is inadequate and useless. In the same manner, in more than two decades, the simple shift-only models have led to no useful algorithm.

Although by limiting the patch size in the stereoscopic (overlapped imagery) correspondence problem, we will not realize a simplification of the pattern differences by shift operations, we do realize some legitimate simplification of pattern differences, thanks to the generally valid "local flatness of surface" assumption: The relationship between corresponding local image patches can be described as an affine transform.

An accurate local pattern matching must be performed by simultaneously find out the shift and skew parameters. The minimum adequate model is the affine Lie group model.

In general, to make geometric measurement, the local pattern differences between binocular or motion stereoscopic images must be parameterized by a few real numbers, such as x-shift, y-shift, rotations, skew transforms, etc. These parameterized pattern differences must be the transforms in a certain Lie group. The Lie group for modeling and parameterizing the local pattern differences in binocular or motion stereo is the 2-D affine group. What the vision system is doing in surface perception is measure local affine parameters, which directly provide orientation and distance information at each location.

The local measurement of the affine parameters of image changes can be performed by an analog process of gradient descent in a simple neural circuit. The dynamical system is driven by "forces" from hypercomplex cells called Lie germs. Lie germs are Lie derivative operators. In the cortex, they are realized as intrinsic neurons which modulate the activity of the principle neuron synapses via their arborized axons.

The massive array of affine sensing circuits functions as a smart sensor tightly coupled to a the conventional passive imaging sensor (retinas). The development of such a massive array requires a very large scale integration (VLSI) of analog circuits. The analog process of the smart sensor can be numerically simulated in a digital computer. With currently available DSP chip technology, the digital simulation engine can be real-time and affordable, depending on processing requirements.

For a long time, in computational vision research, the affine transforms of local visual pattern in stereo imagery, which are crucial to accurately convey 3-D shape information about the surface structure of the visual world, were evaded as "local distortions", as the trouble maker of the favorite shift-and-matching scheme. From the early "zero-crossing" detection3 to later more sophisticated multiple scale feature detections and matching, the rationale behind the efforts of making initial abstracts before matching has been to avoid the difficulties caused by local affine variations of visual patterns.

The problem of dealing with the affine transform effect became unavoidable when the alternate approach of sub-image matching was undertaken. In recent years, it was recognized that to make accurate sub-image matching, the affine transform effect must be taken into account. It was also recognized that these affine transform parameters contain the surface orientation information. There have been effort to formulate algorithms to calculate the image affine flow, instead of the simple optical flow.

The affine flow computation was formulated in image processing domain. Due to the combination of multiple parameters, affine flow is unwieldy to compute with the warping-and-matching, which appears a natural expansion of the old shift-and-matching scheme. Two reasons are: (1) the search space spanned by the combination of multiple affine parameters is tremendous; and (2) warping one image pattern in order to matching the other is much more costly than the simple shift operations. There are efforts to introduce some analytical tools into the computation to make computation less intractable, particularly through the use of derivative computation. However, the image domain is not a proper play ground for use of analytical schemes. Images have no analytical structures at all. They are simply intensity arrays.

Based on the above observation of the shortcomings of the mainstream image understanding effort, we depart from these approaches in two major ways: (1) we consider a Lie group model of local processing of stereoscopic imagery, which means we not only model the image changes with the full 2-D affine transformations, but also furnish it with the parameterization and Lie differentiation mechanisms of a Lie group, and (2) we consider a dual space transform approach, which means we no longer perform image stereoscopic analysis in the image domain, but in its reference frame domain. In particular, we choose several differently oriented Gabor functions as basis functions to build reference frames for local intensity patterns, and affine transform the Gabor basis functions to compensate the image variations due to changed perspectives in motion or binocular stereo. We introduce a principal bundle to the visual field with the 2-D affine Lie group as its local coordinate transformation group.

By introducing Gabor basis functions we are able to perform affine analysis in a very low dimensional space. For example, the dimension of a local image pattern of a 40 by 40 patch is 1,600. If only six differently oriented Gabor functions are taken as a local reference frame, the dimension will be just 6! Operating on Gabor functions, the analytical Lie group structure allows us to fully exploit the advantage of the gradient scheme. The effect of a gradient scheme is to reduce the search space of dimension six to one along a short path towards the minimum energy state.

Thus two economics have been achieved: the high dimension of the vector space of image patches is reduced to the low dimension of their limited (weak) Gabor representation; the 6 dimensions of the parameter space of affine transforms is further reduced to the one dimension of its gradient path.

Thus by introducing the Lie transformation group model, the dual space transform method, and by taking Gabor type receptive field functions as basis functions forming local reference frames, we are able to achieve accurate modeling of image disparity in stereoscopic imagery and substantial reduction in computational complexity. These reductions of dimensionality have effectively turned an intractable computational problem into a practical computational problem.

3 COMPUTATIONAL TEST OF THE MODEL

The core technical issue in 3-D stereoscopic recovery is the "correspondence problem", namely, to determine image geometric (shift and skew) transformation parameters. With these parameters accurately measured, it is straightforward to further determine 3-D structure and motion of visible surfaces.

In a recent benchmark test, we have written a program for numerical simulation of the neural geometric engine process, and calculated affine disparity of a stereoscopic aerial image pair. We have compared our initial results on x-shift with the Defense Mapping Agency's results from photogrammetric measurement of x-shift, in a dense grid (each five pixels is a step in a 1k by 1k aerial stereoscopic image pair, in an overlapped central area of 21,316 pixels total) with accuracy to 0.1 pixel in their x-shift measurement. About 54difference, and has about 5larger than 5 pixels occur at depth edges such as cliffs or steep slopes and valleys, where a small difference in image location results in a large disparity difference but still represents the same 3-D structure of the terrain surface.

The process fails to converge mostly in places with depth discontinuity or deep surface concavity of the surface. At these places, a number 0 is assigned. However, at closely nearby places, where the processes do converge, the numbers are very close to the DMA data at the place. Therefore, the small percentage of places with large differences will not in general affect accurate 3-D terrain reconstruction.

The Neural Geometric Engine (NGE) algorithm measures several affine parameters between patches from a stereoscopic image pair at locations where images are overlapped. These affine parameters, the shift and skew parameters, give both the range and surface orientation information simultaneously. The direct outcome of our NGE algorithm is the so-called 21/2-D representation of scene, namely, a scene described as a surface specified by ranges and surface normals at each visual direction.

Figure 1 shows aerial stereoscopic terrain images of a mountain area in Colorado. Image x-shifts range from a few pixels for deep valley areas to more than 100 pixels for the high peak of the mountain. On the steep slope, image patterns are substantially skewed. The tilt of the airplane also caused up to 5 pixels y-shift. There are large featureless areas, and substantial intensity scale changes between frames. Figure 2 summarizes the result of our computer simulation of the neural dynamical process. Figure 3 shows the positions of excellent convergence (marked by bright spots) are quite densely distributed even nearby the depth edges such as deep cancavities or steep slopes.

The initial benchmark test result has shown that the 3-D recovery of surface shape, even in rough terrain, which is often found in realistic situations, can be achieved very accurately.

4 ANALOG VLSI NEUROMORPHIC COMPUTING DEVICE

An analog VLSI neuromorphic computing device can be designed and fabricated to demonstrate the working principle of the proposed Lie germs and dynamical receptive fields, to prove the feasibility and to provide insight into the robust physical algorithm for affine disparity computation. The proof-of-concept prototype will provide new knowledge for the study of visual motion perception and it will give real-time performance.

Analog VLSI4,5 provides a very attractive way of implementing neural circuitry for mimicking human visual perception. Some of the features of analog VLSI include: (1) Real-time computing capability due to the fact that many primitive arithmetic operations such as addition, subtraction, multiplication, can be implemented by collecting the currents or voltages across the specifically designed analog unit, which can be done within the order of nanoseconds. Higher order computations, such as differentiation and integration, can be derived from the collection of primitives. (2) Capacity for parallel computing in an asynchronous fashion. A properly designed analog computation unit with many simple processing elements can be operated in parallel. Inside each subsystem, there are usually no rigid timing restrictions. This asynchronous computation contributes to simpler timing hardware and faster computation speed. (3) Lower power consumption compared to its digital counterpart. It can be proved that the digital computation for each primitive arithmetic requires orders of magnitude more power to accomplish. For example, digital addition by using full adders requires many state changes of digital gates. Roughly the power necessary for a single digital state change is equivalent to that consumed for performing an analog primitive computation. Hence, analog VLSI implementation consumes much less power than digital implementation. (4) The same functional unit implemented in analog VLSI is much smaller in size as a result of lower power consumption (therefore smaller feature dimension), simpler circuitry, and parallel computation.

The computational primitives of the analog neuromorphic motion perception device 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 simple, complex cells, and Lie germ type hypercomplex cells in visual cortex, as well as the 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 the 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 functions the exponential mapping for transforming the receptive fields, since the receptive fields take the Gaussian distribution function as the basic form of spatial extension. Based on these elemental operations on signals, geometric operations on "receptive fields" can further be built. Figure 4 shows an analog chip which can perform "receptive field" dual scale transform to compensate the scale transform in the intensity pattern, built according to our Lie group model.

5 CONCLUDING SUMMARY

Various new computer architectures have achieved impressive progress in speed and storage. They provided new possibilities in image and signal processing. The neural geometric engine is different from these computers in its 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 analog computing. Moreover, as a neural computing system, it not only 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 processing is different from a computer vision algorithm. It is a digital simulation of the deterministic analog process in a neural circuit, while computer vision algorithms, with feature matching as the central piece, derive from a common sense method of image data processing. The control of our system is the dynamics: specifically, the force of the energy gradient. The control of artificial intelligence vision algorithms are decisions of a logical nature. 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 visual cortex. The spatio-geometric information extracted from the neural geometric engine can be used for various passive sensor based measurements and modelling. It also opens up a new approach to invariant object recognition.

ACKNOWLEDGEMENTS

The development of computational structure of neuromorphic geometric engine and its computer simulation and benchmark test was done by Dr. Thomas Tsao, funded by BMDO SBIR contract DASG60-96-C-0058. The analog integrated circuit chip was designed and tested by Doris Tsao during her undergraduate study at Caltech, in relevant computational neural science courses.

REFERENCES

1. Tsao, T.R. "An Artificial Neural System with Lie Germs for Affine Invariant Pattern Recognition," in: *Proceedings of the SPIE's OE/Aerospace Sensing Orlando '95 Conference*, April, (1995)

2. Tsao, T.R. "A Lie Group Approach to Neural Computation of Image Flow and Binocular Affine Disparity", in: *Proceedings of the SPIE's OE/Aerospace Sensing Orlando '95 Conference*, April, (1995)

3. Marr, D. **Vision**, W. H. Freeman and Company, New York, (1980)

4. Mead, C.A. **Analog VLSI and Neural Systems**, Addison-Wesley Publishing Company, Inc, (1989)

5. Mead, C.A. "Silicon Models of Neural Computation," in: *Proceedings of First International Conference on Neural Network*, San Diego, June 21-24, 1987

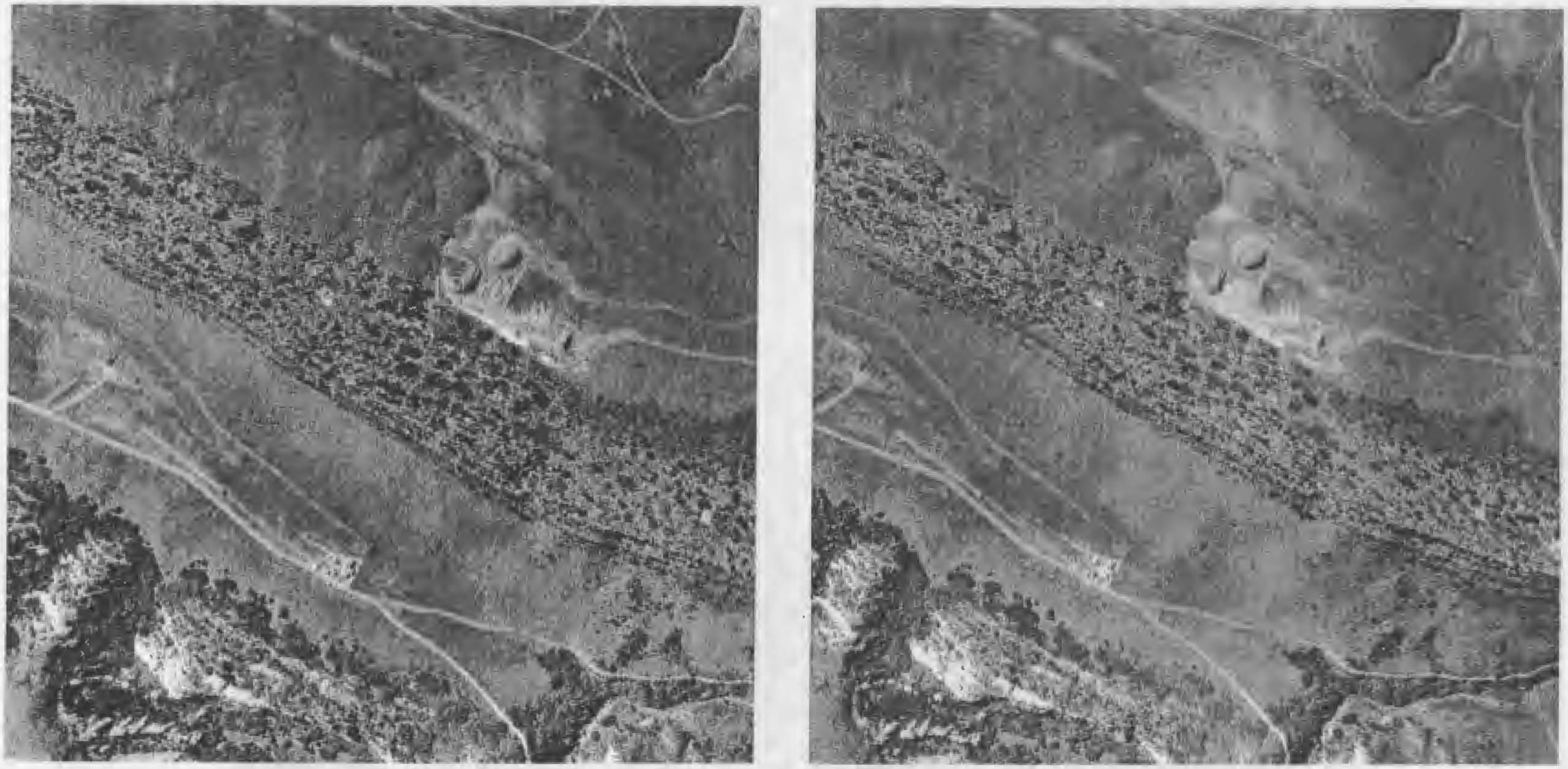

Figure 1. Stereoscopic aerial photos of terrain of a mountain area in Colorado. Pixels has x-shift up to over 100 pixels at the peaks of the mountain and only a few pixels in the deep valley. There are up to 5 pixels y-shift may be caused by tilt of airplane, or may caused by optical distortion.

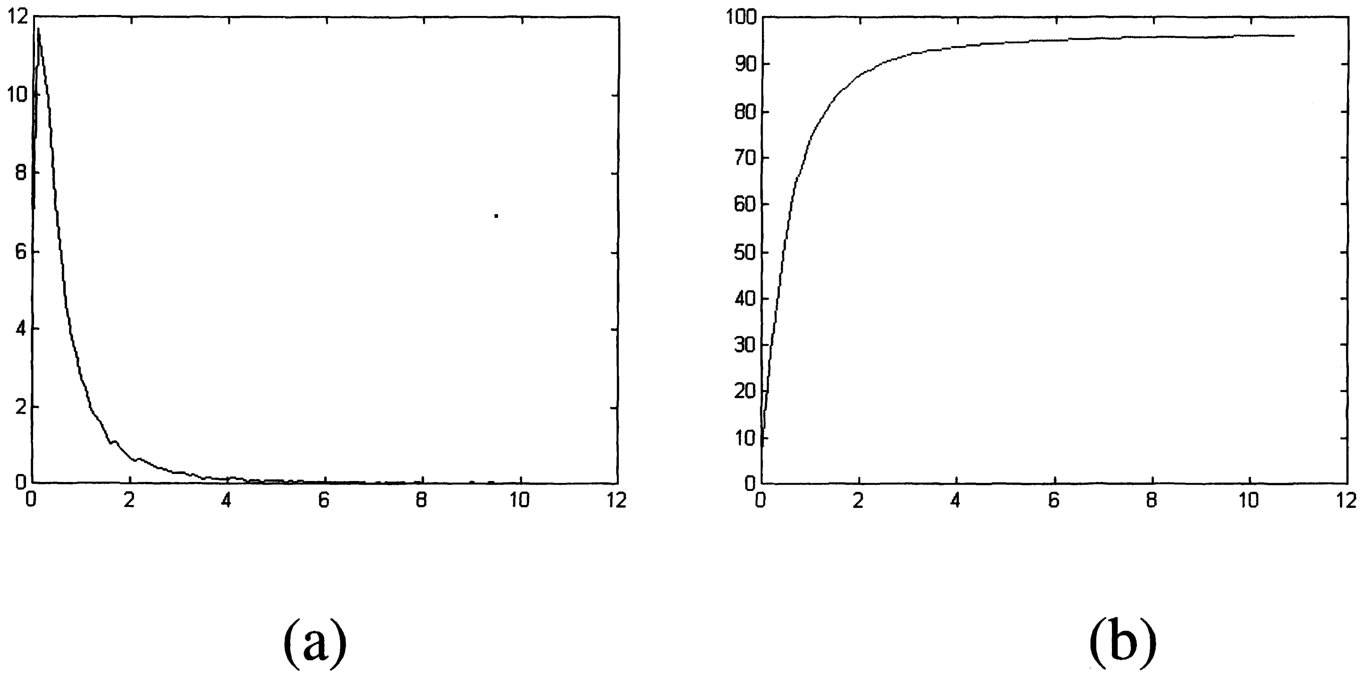

Figure 2. The vertical axis indicates the percent of total pixels, the horizontal axis indicates the difference with DMA "truth" in terms of number of pixels. In (a), the percentages of pixels collected in each 0.1 pixel bucket. The peak is at 0.1 pixel bucket, where the 11.67% of total measurements is fall in. In (b), the accumulated percentage is plotted against the distance to DMA truth in terms of pixels. 90% of total measurements is within 2.5 pixel distance. After five pixel distance, which 95% measurements are in, mostly the left are not convergent positions.

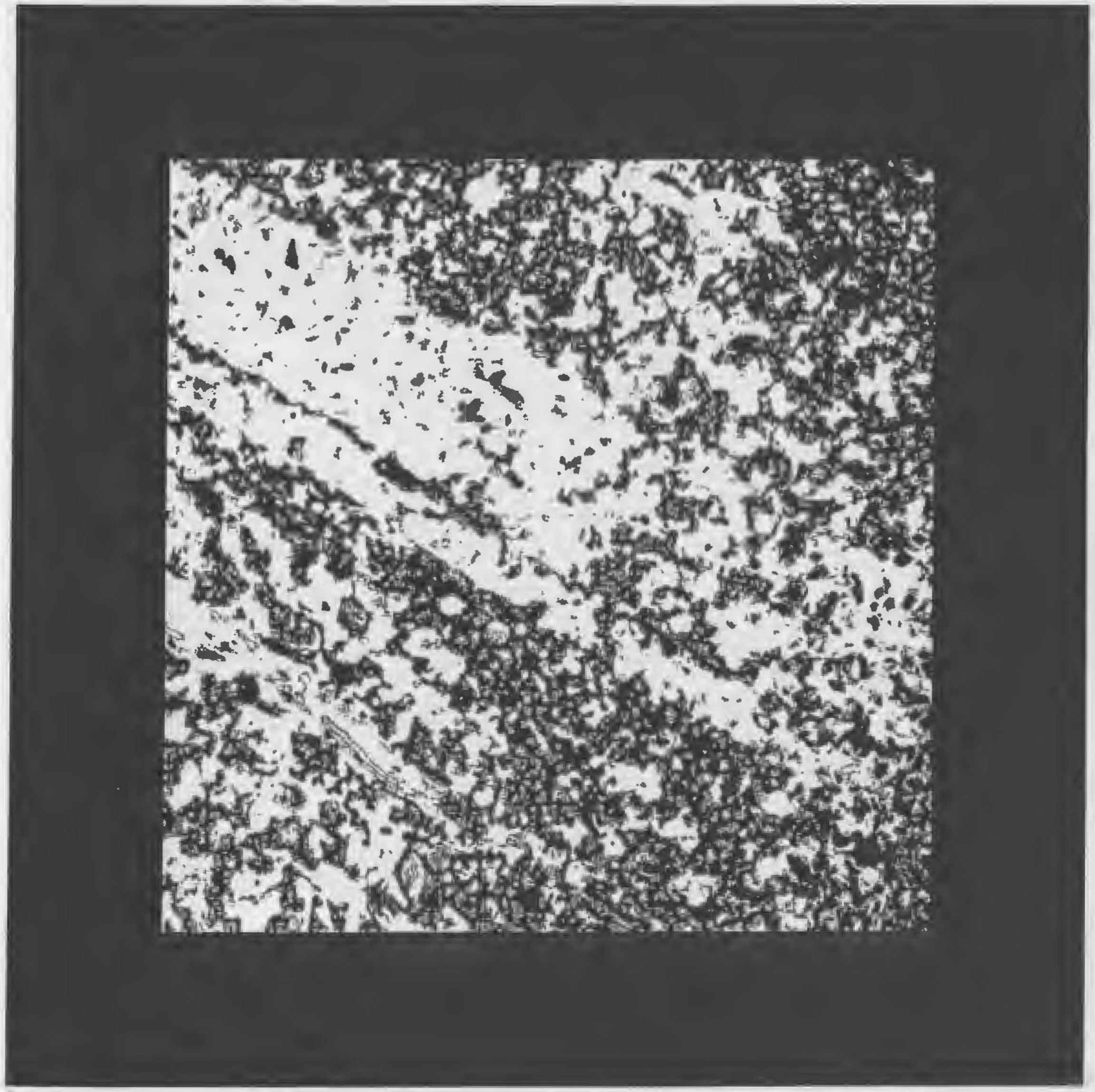

Figure 3. Three gray levels are used to indicate convergence status of binocular fusion processes: dark spots are those where the initial run of dynamical process not converge; gray spots are those where the energy reduced to less than one third of reference level; and white spots are those where the energy rediced to less than one tenth of reference level.

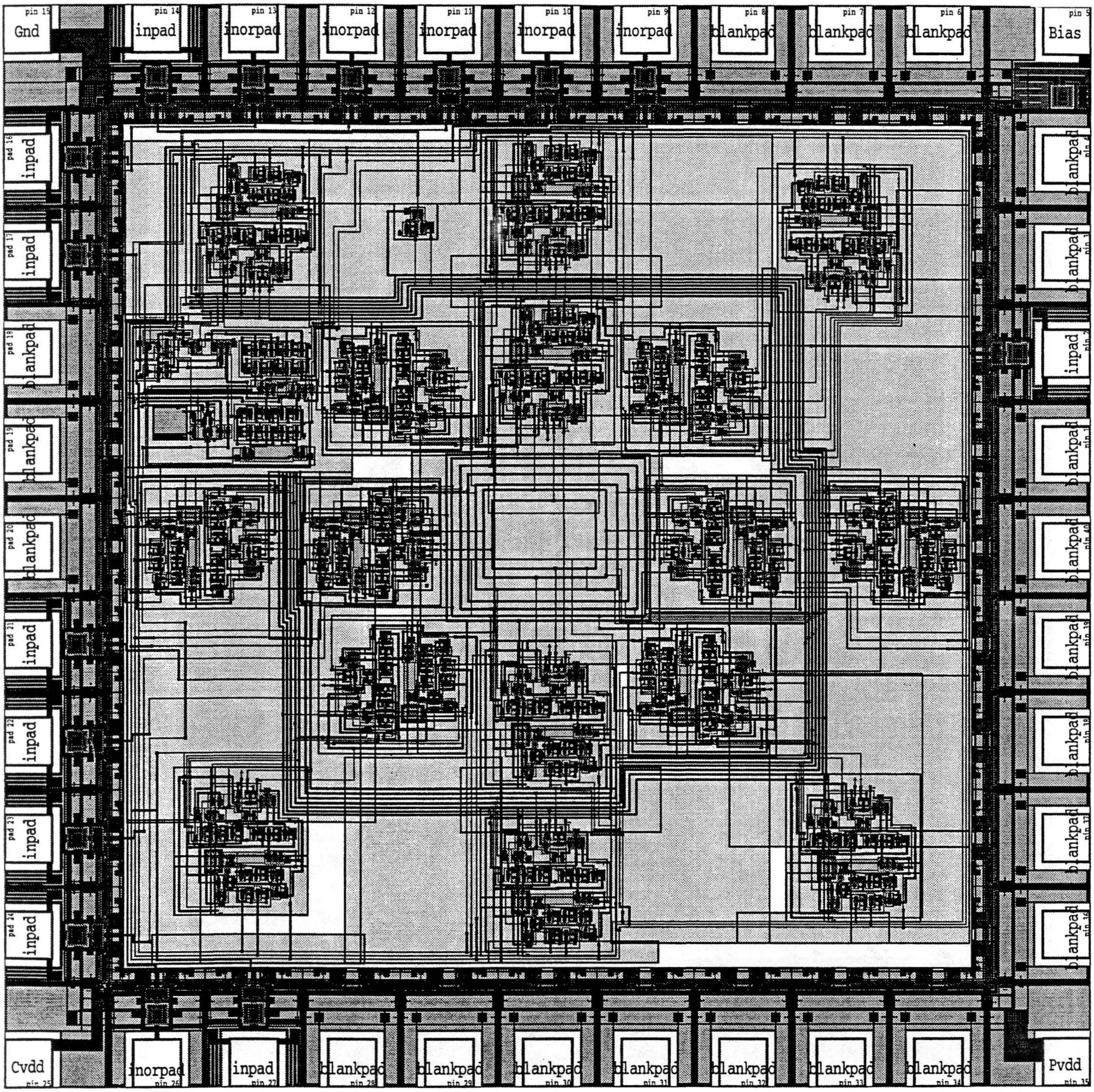

Figure 4. Shows the chip layout of an analog vision chip (designed by Doris Tsao for Analog VLSI design class at CalTech) which perform pattern matching with the Lie group dynamical receptive field method. The receptive field of this chip is composed of 16 pixels, arranged roughly in 2 concentric rings.