REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	Nov.25,1996	Final Technical 01/01/93 to 06/30/96

4. TITLE AND SUBTITLE
Computing with Stochastic Signals

5. FUNDING NUMBERS
Contract No. AFOSR F 49620-93-1-0109

AFOSR-TR-97 0423

6. AUTHOR(S)
Christoph von der Malsburg

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Southern California
Computer Science Department

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Dr. Arge Nachman
Electronic and Material Sciences, Bldg 410, Room 213
Bolling AFB Washington, DC 20332-6448
Tel (202) 767-5027, -4939

NM

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release: distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The project had set out to create a computer vision system for the analysis of natural scenes. The system was to split the process into several time scales such that it could later be implemented in parallel hardware (e.g., pixel-parallel, feature-sequential). The PI has fully succeeded in reaching these goals. He has created, tested and published a system that is able to recognize objects (e.g., faces) from freely taken digital images. Recognition is based on stored sample images. Recognition of new object types is possible simply on the basis of new sample images. The system has been extensively tested and optimized, with face recognition from large galleries as test environment. Results have been published. The system is ready for implementation in parallel hardware, such as that developed in the VIGILANTE project by S. Suddarth of BMDO in collaboration with JPL. Due to reduction of funding in the third year the PI has not been able to integrate figure-ground separation in the system.

14. SUBJECT TERMS
Visual object recognition. Invariance to position, robustness to rotation in depth, to illumination, to deformation and to partial occlusion. Learning from single examples. Potential for massively parallel implementation.

15. NUMBER OF PAGES
7+1

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std Z39-18
298-102

19971007 042

Memo to: VON DER MALSBURG, CH

From: Ofelia Galvan
Department of Contracts and Grants, Ext. 07762

Date: 07/01/96

Subject: Report of Inventions - Acc. No. 53-4509-4802

In order to comply with the reporting requirements of the following agreement, we must submit a report of inventions conceived or first actually reduced to practice in the performance of work during this agreement.

Project Title: Computing With Stochastic Signals

Sponsor Award No. F49620-93-1-0109

Sponsor: Air Force Office of Scientific Research

Reporting Period: 01/01/93 THROUGH 06/30/96

Inventions Conceived/Reduced to Practice? ✓ NO or ______ YES

If yes, please identify.

C. von der Malsburg Nov. 25, 1996

Principal Investigator Date

Failure to disclose an invention promptly may result in the loss of important rights including the University's right to retain title to the invention.

PLEASE RETURN A COPY OF THIS MEMO WITH YOUR NOTATION AND SIGNATURE

TO: Ofelia Galvan, STO 330 mc-1147

FINAL REPORT
Award No.: AFOSR F 49620-93-1-0109
Title of Project: "Computing with Stochastic Signals"
Period: 01/01/93 to 06/30/96

Principal Investigator: Christoph von der Malsburg
Computer Science Dept., Univ. of Southern California, Los Angeles

Objectives

The following objectives were pursued in the project:

- Object recognition from digital camera images.
- Invariance with respect to position.
- Robustness with respect to distortion (rotation in depth, object deformation).
- Robustness with respect to illumination, noise, partial occlusion and changing background.
- Preparation for massively parallel implementation.

Our system

Our system is fully described in an article that is accepted for publication [3]. See also the web book [2].

Our system for object recognition and scene analysis is based in an essential way on the Morlet wavelet transformation. This is a series of convolutions, performed on input images. As convolution kernels, Morlet wavelets (also called Gabor-type wavelets) are used. These form a resolution pyramid. A single image point is then characterized by the vector of responses of all kernels centered at that point. This vector is called a jet. A full array of jets, one for each pixel, forms an "image domain."

Objects to be recognized are stored as two-dimensional arrays of such jets which are extracted from the aspect of a sample image of the object. Different objects (and if necessary different aspects of one object) are stored side by side in a "model domain."

An object appearing in the image domain is recognized by graph matching. This is implemented with the help of a full matrix of connections ("dynamic links") between image points and all nodes in all objects. (Although this full matrix is conceptually simple, a sparse matrix is used in actual applications.) This matrix is reorganized as a dynamical system (see Table 1 in [3]). Nodes in a model and nodes (pixels) in the image are compared in terms of their jets. Jet similarity is computed as a scalar product of normalized jets. Matrix reorganization is attempting to install one-to-one connections between nodes in image and model, winning connections tend to have maximal jet similarity, and neighboring nodes tend to be connected to neigboring nodes. Thus, a (distorted) topological connection pattern is set up between image and model(s). Different models compete for connections, and the best-fitting model wins out in a winner-take-all scheme. All these tendencies are implemented in a simple set of dynamical equations.

To handle different aspects of the same object (the object seen under different angle), and to handle variations in one object type, multiple models are stored ("bunch graph," not part of this project). Several such models for one object can be compacted economically into a "fusion graph." This aspect of our system has been implemented and perfected in other projects. Complex scenes, and scenes with heavy object occlusion, necessitate a separate process of figure-ground separation. This can be homogeneously integrated into the system developed here (and it was part of the original project), but due to severe budget cuts in the third year of this project that function was not implemented here (although we have demonstrated the figure-ground separation process in another project).

In separate work we have reduced the dynamic link matrix to a low-dimensional entity that can be efficiently and quickly optimized on a digital machine. We have been able to achieve real-time object recognition from video input (not part of this project).

To a large part, the system developed in this project has been shaped by considerations for massively parallel implementation. In one extreme case, the wavelet transformation can be implemented by a fully optical system that is pixel-parallel and wavelet-sequential. Jet similarity can then be implemented as a temporal correlation between local signals. With appropriate photorefractive materials the dynamic link matrix can be efficiently implemented. Further development is needed to realize this potential. In a less advanced system, electronic analog signal processing architectures or DSP arrays can be used to compute the wavelet transform in real time at high frame rates and to process jet similarities at extremely high rates. The company Siemens is developing a vision chip (SEE-1) that is specifically targeting our application. That chip will be centered around an array of 128 DSPs, will have a sustained computing power of 5 GFLOP, and will enable us to analyze complex video sequences in real time.

Accomplishments/New Findings

We have applied our system to the recognition of objects (human faces) against large data bases (more than 100 persons) with only one image per person. We have achieved correct recognition with high reliability (in more than 80% of the cases the correct gallery face was ranked as number one) in spite of rotation in depth of up to 15 degrees and changes in facial expression (see [3]). With depth-rotation of 30°, the recognition rate fell to 66% correct identification, a rate still comparable to that of human subjects.

We have been able to decisively improve the performance of our system with the help of several important changes to its dynamics:

- Conspiracy of corresponding parts of stored images — see second term on the right-hand-side of the h-equation of Table 1 in [3].
- Replacement of additive signal combination by maximal signal (5th term h-equation, Table 1 of [3]).

- Introduction of an "attention"-variable to manage the region-of-interest in the image — a-equation, Table 1 in [3].

We have recently made the important discovery that object recognition in our system can be speeded up by a very large factor by forcing decision between the various stored models without re-organization of the Image-Model matrix. According to our experiments, the correct recognition reliability is then almost as high as with full matrix re-organization.

Our system has several features that let it stand out compared to other object recognition systems:

- No learning or training is required (although system performance can be improved over the present level by training).
- Flexibility/generality with respect to object types: no manual programming or construction of object models is required. The system is based on the simple storage of images.
- Potential for object recognition from any angle or for better generalization over object variation by storing (and consolidating) more views.
- In relation to conventional optical correlation methods: ability to handle distortion, partial occlusion and changing background.
- In distinction to other image understanding systems, ours is very homogeneous, the computing intensive parts having the form of convolutions and scalar products of vectors. Our system thus supports easy and economical implementation in parallel hardware.

Due to its flexibility, our system has the potential for important applications in the military and civilian domain. Among these are:

Military applications:

- Target recognition

- Battle scene analysis
- Security access systems

Civilian applications:

- Area surveillance,
- Security access systems (in fact, we are already engaged in commercial application of a related system)
- Manufactury (quality control, object manipulation)

Personnel

Dr. Michael Lyons (postdoc); Laurenz Wiskott (Research Assistant).

Publications

[1] L. Wiskott and C. von der Malsburg: Recognizing Faces by Dynamic Link Matching. In: Proceedings of the International Conference on Artificial Neural Networks, Paris 1995, pp. 347–352 (refereed contributed paper).

[2] L. Wiskott and C. von der Malsburg: Face Recognition by Dynamic Link Matching. In: Lateral Interactions in the Cortex: Structure and Function. Electronic book, Sirosh, J. and Miikkulainen, R. and Y. Choe (editors), chapter 4, 1995. (ISBN 0-9647060-0-8) http://www.cs.utexas.edu/users/nn/web-pubs/htmlbook96/.

[3] L. Wiskott and C. von der Malsburg: Face Recognition by Dynamic Link Matching. Neural Computation. (Accepted for publication).

Interactions/Transitions

Consultative and advisory functions

None

Transitions

None

New discoveries, inventions, or patent disclosures

None

Honors/Awards

Pioneer Award 1994 of the Neural Network Councel of the IEEE.