

AD-A233 019

DOCUMENTATION PAGE



1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE MAR 27 1991			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
4. PERFORMING ORGANIZATION REPORT NUMBER(S)			7a. NAME OF MONITORING ORGANIZATION Cognitive Science Program Office of Naval Research (Code 1142PT)		
6a. NAME OF PERFORMING ORGANIZATION The Regents of the University of California		6b. OFFICE SYMBOL (if applicable)	7b. ADDRESS (City, State, and ZIP Code) 800 North Quincy Street Arlington, VA 22217-5000		
6c. ADDRESS (City, State, and ZIP Code) University of California, Los Angeles Office of Contracts and Grants Administration Los Angeles, California 90024		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0395			
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Defense Advanced Research Projects Agency		8b. OFFICE SYMBOL (if applicable)	10. SOURCE OF FUNDING NUMBERS		
8c. ADDRESS (City, State, and ZIP Code) 1400 Wilson Boulevard Arlington, VA 22209-2308		PROGRAM ELEMENT NO. 61153N	PROJECT NO. RR04206	TASK NO. RR04206-OC	WORK UNIT ACCESSION NO. 442c022
11. TITLE (Include Security Classification) Artificial Intelligence Measurement System (Briefing Charts)					
12. PERSONAL AUTHOR(S) Baker, Eva L.					
13a. TYPE OF REPORT Interim		13b. TIME COVERED FROM 7/1/86 TO 3/31/87		14. DATE OF REPORT (Year, Month, Day) March 1987	
15. PAGE COUNT 63					
16. SUPPLEMENTARY NOTATION Briefing charts used at the ONR Contractors' Meeting, Yale University, March 1987.					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP	Artificial intelligence, natural language understanding, expert systems, expert system shells, human benchmarking, machine vision		
12	05				
05	07				
19. ABSTRACT (Continue on reverse if necessary and identify by block number)					
These briefing charts provide an overview of the research plan for the Artificial Intelligence Measurement System (AIMS). They provide the goals and proposed implementation procedures for the major areas of inquiry--natural language understanding, machine vision, expert systems (including expert system shells), technology assessment.					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Susan Chipman			22b. TELEPHONE (Include Area Code) (703) 696-4318		22c. OFFICE SYMBOL ONR 1142CS

Project Report #1

ARTIFICIAL INTELLIGENCE MEASUREMENT SYSTEM
(Briefing Charts)

Eva L. Baker

Center for Technology Assessment
UCLA Center for the Study of Evaluation

These briefing charts were used at the ONR Contractors' Meeting, Yale University, March 1987.

Artificial Intelligence Measurement System
Contract Number N00014-86-K-0395

Principal Investigator: Eva L. Baker

Center for Technology Assessment
UCLA Center for the Study of Evaluation

Approved For	
THIS CHART	
FOR THE	
U.S. GOVERNMENT	
JANUARY 1987	
By	
Distribution	
Availability Codes	
Dist	AVAIL & USE
A-1	SECRET

This research report was supported by contract number N00014-86-K-0395 from the Defense Advanced Research Projects Agency (DARPA), administered by the Office of Naval Research (ONR), to the UCLA Center for the Study of Evaluation. However, the opinions expressed do not necessarily reflect the positions of DARPA or ONR, and no official endorsement by either organization should be inferred. Reproduction in whole or part is permitted for any purpose of the United States Government.

Approved for public release; distribution unlimited.

SPONSORS

**OFFICE OF NAVAL RESEARCH
SUSAN CHIPMAN**

**DEFENSE ADVANCED RESEARCH PROJECTS
AGENCY
STEVE KAISLER**

ARTIFICIAL INTELLIGENCE MEASUREMENT

SYSTEM

(AIMS)

EVA BAKER, P.I.

MARCH 1987

**ONR CONTRACTORS' MEETING
YALE UNIVERSITY**

BACKGROUND

ICAI FORMATIVE EVALUATION

ARI-JPL 1983-85

WEST

PROUST

LESSONS LEARNED

- 1. LIGHT FOCUS ON EFFECTIVENESS IN ICAI COMMUNITY
(PROUST IS AN EXCEPTION)**
- 2. EVEN AI APPLICATIONS DEVELOPERS FOCUS ON
THEORY BUILDING RATHER THAN OUTCOMES**
- 3. NEED FOR AIMS**

UCLA

**CENTER FOR THE STUDY OF EVALUATION
DEPARTMENT OF COMPUTER SCIENCE**

HELP FROM:

EDUCATIONAL TESTING SERVICE

AIR FORCE HUMAN RESOURCES LABORATORY

NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER

NAVAL OCEAN SYSTEMS CENTER

PERCEPTRONICS

UNIVERSITY OF ILLINOIS

CONSULTANTS ROUND THE WORLD

PROJECT SLOGAN:

HOW SMART ARE SMART COMPUTER SYSTEMS?

GOAL:

- **TO DESCRIBE THE EFFECTIVENESS OF AI APPLICATIONS IN MULTIPLE TERMS**
- **TO DEVELOP AND TEST APPROACHES TO USING HUMAN BENCHMARKS**

- **AVOID OVERLY SIMPLIFIED ANSWERS**
- **WE ARE NOT LOOKING FOR A SINGLE
NUMBER OR SINGLE CRITERION**

PROJECT INSPIRATIONAL MESSAGE:

**ASK NOT WHAT COMPUTER SCIENCE CAN DO
FOR MEASUREMENT . . .**

APPROACH

MULTIPLE CHARACTERIZATION OF EFFECTIVENESS OF AI IMPLEMENTATIONS

- 1. PLACE IN THE SCHEME OF THINGS**
- 2. PROCESS ANALYSIS**
- 3. OUTCOMES COMPARED TO PEOPLE ON TESTBED
DOMAINS**
- 4. COMPARATIVE SCALES**

**COMMON TECHNICAL APPROACH FOR NATURAL
LANGUAGE AND VISION***

**SEPARATE APPROACHES FOR EXPERT
SYSTEMS SHELLS AND TECHNOLOGY
ASSESSMENT**

***WITH IMPORTANT DIFFERENCES TO BE
DESCRIBED ANON**

RESEARCH AREAS

NATURAL LANGUAGE UNDERSTANDING

VISION

EXPERT SYSTEM SHELLS

APPROACHES TO TECHNOLOGY ASSESSMENT

CONCEPTUAL ANALYSIS APPROACH*

- **MAP AREA**
- **REVIEW AND AMEND FRAMEWORK FOR MAP**
- **FILL IN EXAMPLES**
- **REVIEW IN TERMS OF HUMAN PERFORMANCE**
- **ANALYZE DIFFERENCES, AREAS FOR PROGRESS, RESEARCH FUTURES, ETC.**

***ORDER OF STEPS WILL DIFFER FOR NATURAL LANGUAGE AND VISION**

COMMON TECHNICAL APPROACH IN NATURAL LANGUAGE AND VISION

- 1. CONCEPTUAL ANALYSIS OF AREA**
- 2. DEVELOP AND REFINE PROCESS CRITERIA FOR ASSESSING IMPLEMENTATIONS**
- 3. REFINEMENT OF BENCHMARKING MODEL**
- 4. CONDUCT STUDY OF PARTICULAR IMPLEMENTATIONS**

BENCHMARK APPROACH

**TO DESCRIBE IMPLEMENTATION'S EFFECTIVENESS IN TERMS
OF EMPIRICAL HUMAN PERFORMANCE MEASUREMENT**

- **DEPENDS UPON USING A DOMAIN REFERENCED
ACHIEVEMENT TEST MODEL**
- **DEPENDS UPON CHARACTERIZING COMPUTER
PERFORMANCE IN TERMS OF AN IDENTIFIABLE
POPULATION, E.G., 4TH GRADE KIDS, PEOPLE
WITH A 50% SIGHT LOSS**

PROCESS ANALYSIS CRITERIA

(HIGHLIGHTS THEREOF)

- **DESIGN GOALS AND FUNCTIONAL SPECIFICATION**
(PROBLEMS, APPROPRIATENESS, INTERFACE, CONSTRAINTS)
- **COGNITIVE PROCESS MODEL**
(ASSUMPTIONS, SUFFICIENCY, PARSIMONY, RESEARCH BASE)
- **IMPLEMENTATION**
(PARTS COMPLETED, LEVEL, ALGORITHMS, EASE OF EVALUATION AND MAINTENANCE, GENERALIZABILITY, ADAPTABILITY, HARDWARE, SOFTWARE REQUIREMENTS)

PERFORMANCE BENCHMARKING

- **TASKS SOLVED BY COMPUTER ANALYZED INTO PARTICULAR DOMAINS**
- **SPECIFICATIONS FOR "TEST" ITEMS CREATED BASED UPON THESE DOMAINS**
- **"ITEMS" FOUND (ETS) OR CREATED**
- **PEOPLE BRACKET TESTED UNTIL APPROPRIATE POPULATIONS FOUND**
- **IMPLEMENTATION CHARACTERIZED IN TERMS OF POPULATION PERFORMANCE**

HOW HUMAN PERFORMANCE BENCHMARKING WILL WORK

HARD STUFF

- **DEALING WITH CONSTRICTED DOMAINS
(RELIABILITY)**
- **FINDING A COMMON BASIS FOR SCALING
ACROSS DOMAINS OR AI IMPLEMENTA-
TIONS WITHIN AREA**

OVERVIEW OF THE NATURAL LANGUAGE AREA

NATURAL LANGUAGE SOURCEBOOK AND DATABASE

- **ONLINE DATABASE QUERY/RETRIEVAL SYSTEM**
- **HARDCOPY VERSION OF DATABASE**
- **REFERENCE SOURCE OF ISSUES AND PROBLEMS IN
COMPUTATIONAL ANALYSIS OF NATURAL LANGUAGE**

NATURAL LANGUAGE CONCEPTUAL ANALYSIS

TO BE BASED ON A SOURCEBOOK (BOTTOM-UP)

. . . DYER'S GROUP

PROCESS ANALYSES OF IMPLEMENTATIONS

. . . DYER'S GROUP PLUS HELP

HUMAN BENCHMARKS

. . . BAKER'S GROUP

USES OF NATURAL LANGUAGE SOURCEBOOK/DATABASE

- **REFERENCE SOURCE FOR NL RESEARCHERS,
STUDENTS, USERS**
- **MECHANISM FOR AUTOMATICALLY TESTING LANGUAGE
ANALYZERS**
- **BENCHMARK CRITERIA FOR MEASURING CAPABILITIES
OF LANGUAGE PROCESSING SYSTEMS**

CONTENTS OF NATURAL LANGUAGE SOURCEBOOK/DATABASE

- **COMPENDIUM OF EXEMPLARS WHICH HAVE RAISED
QUESTIONS FOR AI/NLP RESEARCHERS RESEARCHERS
AND LINGUISTS**
- **DISCUSSION OF ISSUES RAISED BY EACH EXEMPLAR**
- **BIBLIOGRAPHIC REFERENCES FOR EACH EXEMPLAR**
- **CROSS REFERENCES TO ANALYZERS WHICH CLAIM TO
DEAL WITH EXAMPLES**

TECHNICAL APPROACH:

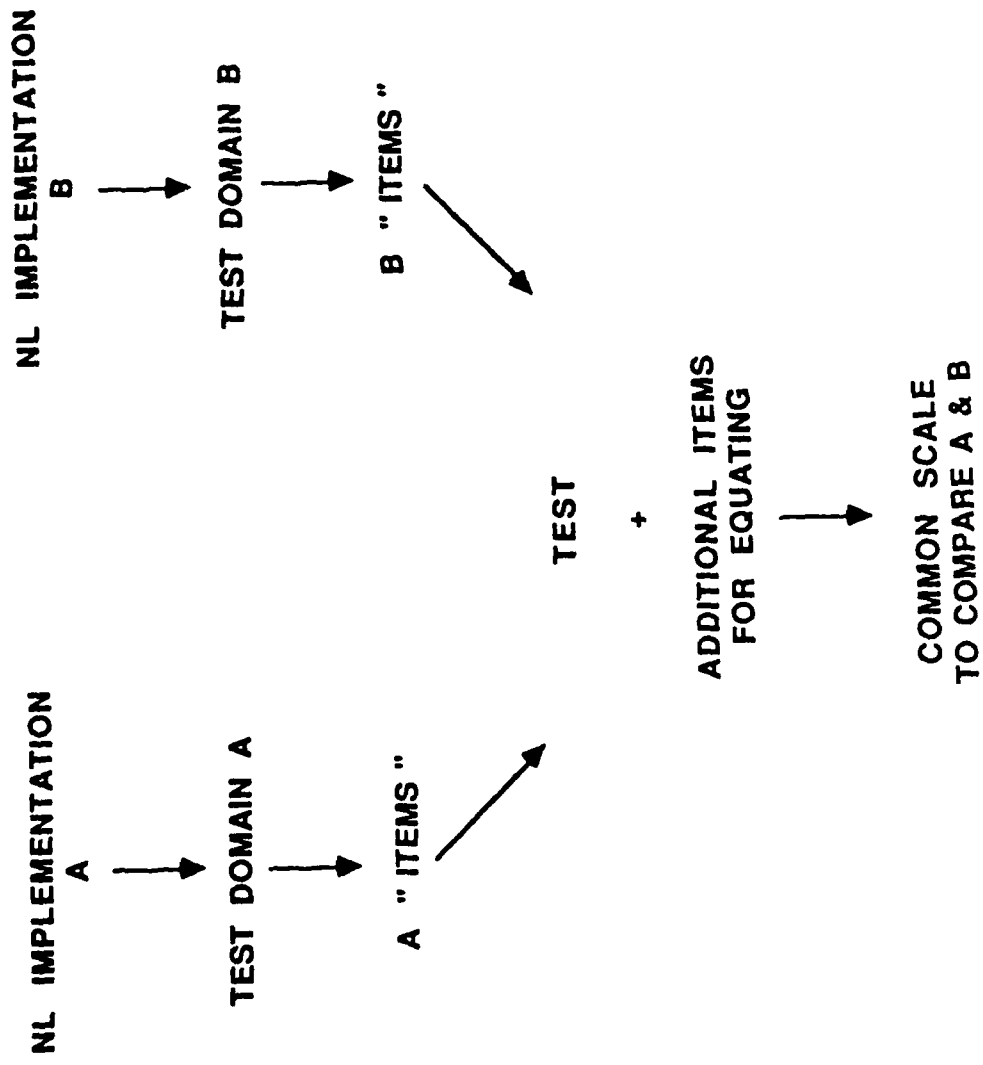
- **THE PROCESS ANALYSIS OF IRUS**
- **IRUS AS IMPLEMENTED IN A PARTICULAR DOMAIN**

**NATURAL LANGUAGE IMPLEMENTATION TO BE
STUDIED FOR PROCESS ANALYSES AND
BENCHMARKING**

IRUS

**IRUS STANDS FOR INFORMATION RETRIEVAL
USING THE RUS PARSER**

BENCHMARK COMPARISON EXAMPLE



BENEFITS OF IRUS

- **REPRESENTS COMMON APPLICATION:
NL FRONT-END TO A DATABASE**
- **HELP FROM THE GOVERNMENT IN
GETTING ACCESS**

IRUS QUERY EXAMPLES

- **HOW MANY PEOPLE IN DEPT. 45 DON'T
LIVE IN CAMBRIDGE OR BOSTON? LIST
THEIR NAMES.**
- **HAS EVERY WOMAN IN DEPT. 45 WRITTEN
A PAPER SINCE 1979?**

(FROM BATES, MOSTER, STALLARD, BBN, 1984)

PROBLEMS WITH IRUS

- **FRONT-END TO A DATABASE**
- **ASSESSMENT MUST DISENTANGLE
UNDERSTANDING FROM DATABASE
FUNCTIONS**
- **IRUS IS ABOUT TO BE REDONE**

IRUS IN A PARTICULAR DOMAIN

FRESH

**INTERESTING FACTS ABOUT FRESH
INFLUENCING OUR ANALYSES**

FRESH USES NL MENU

FRESH IS ALSO TOP SECRET

**IRUS IN THE FRESH CONTEXT WILL BE
ASSESSED USING HELP FROM**

NOSC

NPRDC

AFHRL

EXPERT SYSTEM SHELL AREA

- **EARLY PLANNING STAGE**
- **EVALUATION QUESTIONS**
- **TECHNICAL APPROACH**

PRELIMINARY EVALUATION QUESTIONS FOR EXPERT SYSTEM SHELLS

- 1. HOW USEFUL IS THE TOOL IN THIS DOMAIN?**
(FACILITY, TIME, EXPERTISE, EFFICIENCY)
- 2. HOW WELL DOES THE TOOL FACILITATE THE INTEGRATION
OF NEW KNOWLEDGE IN THE DOMAIN?**
(ADD, AMEND, REFINE, CONSISTENCY)
- 3. HOW WELL SUITED IS THE KNOWLEDGE REPRESENTATION
FOR TESTBED DOMAINS?**
(ADAPTABILITY, UNDERSTANDABLE, COMPLETNESS, CONCISION)
- 4. GIVEN REQUIREMENTS, CONSTRAINTS, OUTPUT, HOW DO
CURVES OF PERFORMANCE VARY?**
(BY APPLICATION, EXPERTISE, ETC.)

TECHNICAL APPROACH FOR EXPERT SYSTEM SHELL AREA

**DETAILED CASE STUDIES OF EXISTING SHELLS ON STANDARD
TESTBED PROBLEMS**

PROTOCOL ANALYSES

OUTCOMES:

- **CREATION OF TESTBED PROBLEMS**
- **ANALYSES OF ESS IN TWO TOPIC AREAS**
- **REPORTS OF TIME, COST, AND EFFECTIVENESS OF THE PARTICULAR
SHELLS EXAMINED**
- **EXAMINATION OF UTILITY OF TESTBED PROBLEMS FOR FUTURE
BENCHMARKING**
- **DECISION TO EXTEND APPROACH TO LARGER SCALE EXPERIMENTS**

WELL SPECIFIED
DOMAIN (IRT)

FUZZY DOMAIN
(e.g. HISTORY —
CIVIL WAR SPEECHES)

HIGH POWER SHELL
e.g., ART

MICRO-BASED SHELL(S)
e.g., M-1

PROTOCOL ANALYSIS FOR EXPERT SYSTEM SHELL AREA

- TO DETERMINE RATE AND PROGRESS ON PROTOTYPE APPLICATION**
- TO DETERMINE ENABLING AND DISABLING FACTORS AT EACH STAGE**
- TO DETERMINE FUNCTIONAL TASK REQUIREMENTS**
- USING OBSERVATION (VIDEO) AND INTERVIEW**

**SPIN-OFF USE OF ESS TESTBED DOMAINS
(IRT + SPEECHES)**

- **AI TEST DEVELOPER**
- **DEVELOPMENT OF MEASURES FOR
"DEEP UNDERSTANDING" TASKS**
- **SUPPORTED BY OERI**

TECHNOLOGY ASSESSMENT AREA

QUESTIONS:

WHAT APPROACHES IN COMMON?

WHAT LESSONS LEARNED?

REQUIREMENT:

**COMMON DOCUMENTATION ACROSS ALL THE THREE
PROBLEM AREAS (NL, VISION, ESS)**

CONFERENCE:

PRODUCTS

- 1. CONCEPTUAL ANALYSES (OR MAPS) OF HUMAN PERFORMANCE TASKS AND COMPUTER TASKS FOR NLP AND VISION**
- 2. REPORTS OF SPECIFIC IMPLEMENTATIONS IN NLP AND VISION**
- 3. MODEL FOR HUMAN BENCHMARKING IN APPLICATION AREAS WITH SPECIFIC EXAMPLES IN NLP AND VISION**
- 4. EXPERT SYSTEM SHELL CASE STUDIES**
- 5. LESSONS LEARNED DOCUMENT ON TECHNOLOGY ASSESSMENT**

CSE STAFF ON THE AIMS PROJECT

**EVA BAKER
ELAINE LINDHEIM
RICH SHAVELSON
MERL WITTROCK
BENGT MUTHEN
DEAN SLAWSON**

NATURAL LANGUAGE

MICHAEL DYER

ALEX QUILICI

JOHN REEVES

RIC FEIFER

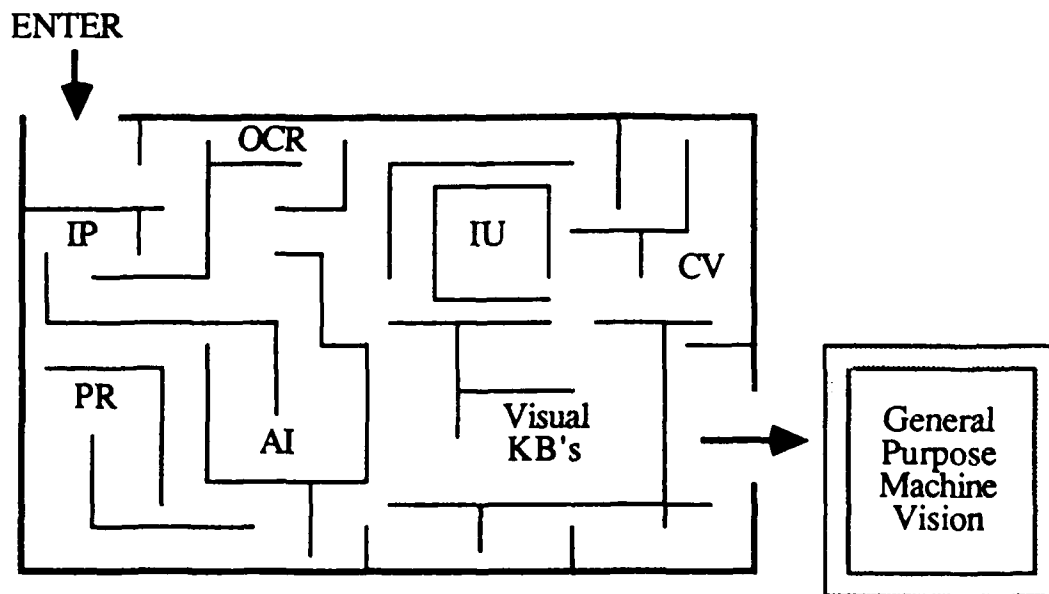
VISION

**JOSEF SKRZYPEK
DAVID GUNGNER
EDMOND MESROBIAN
EUGENE PAIK
MICHAEL STIBER**

AND A CAST OF THOUSANDS AS CONSULTANTS

**MANAGEMENT HELP: JOAN HERMAN
 KATHARINE FRY**

WHICH WAY DO WE GO?



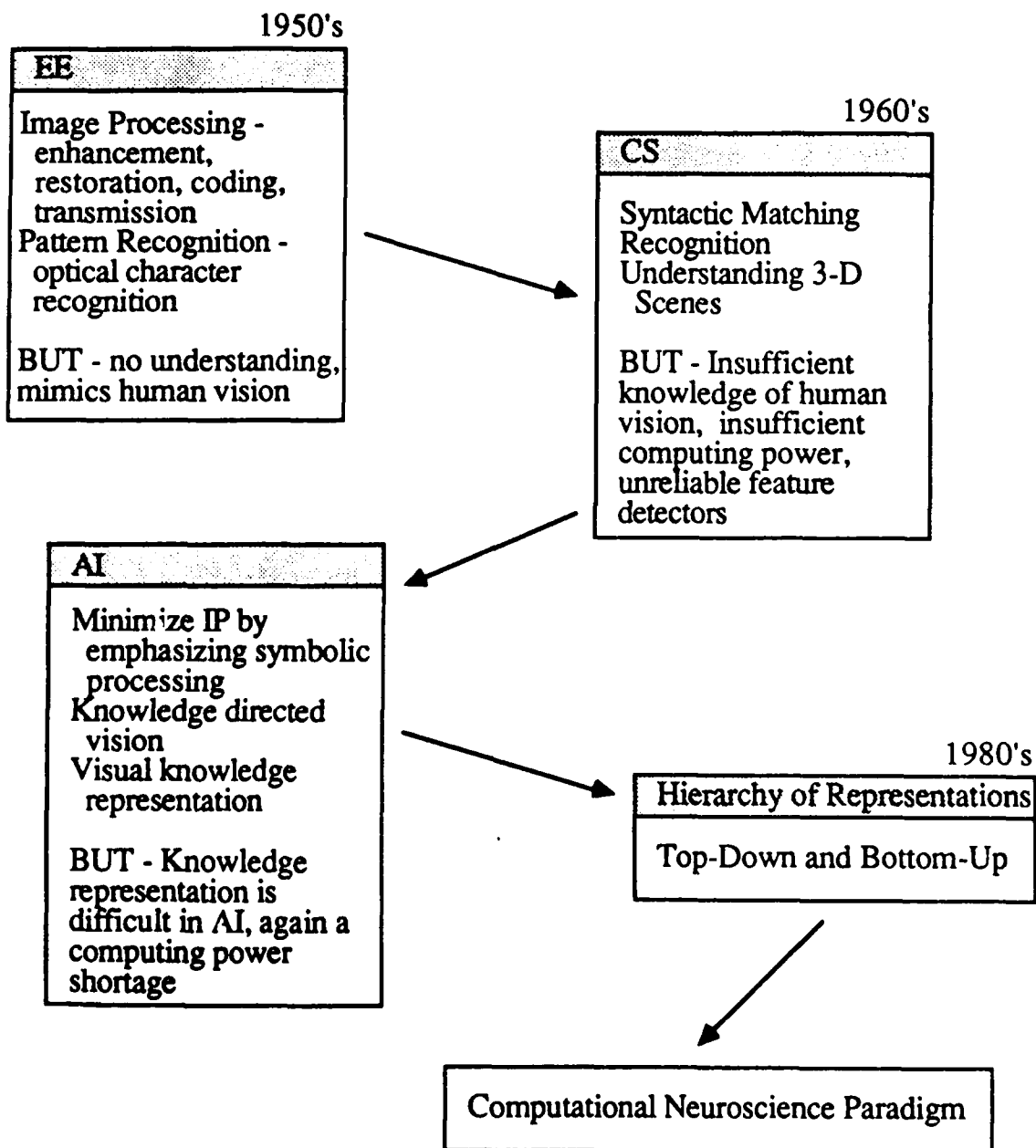
CATEGORIES OF PROCESSING

Processing {
VP = Video Processing
IP = Image Processing
PP = Picture Processing

Classification {
PR = Pattern Recognition
II = Image Interpretation

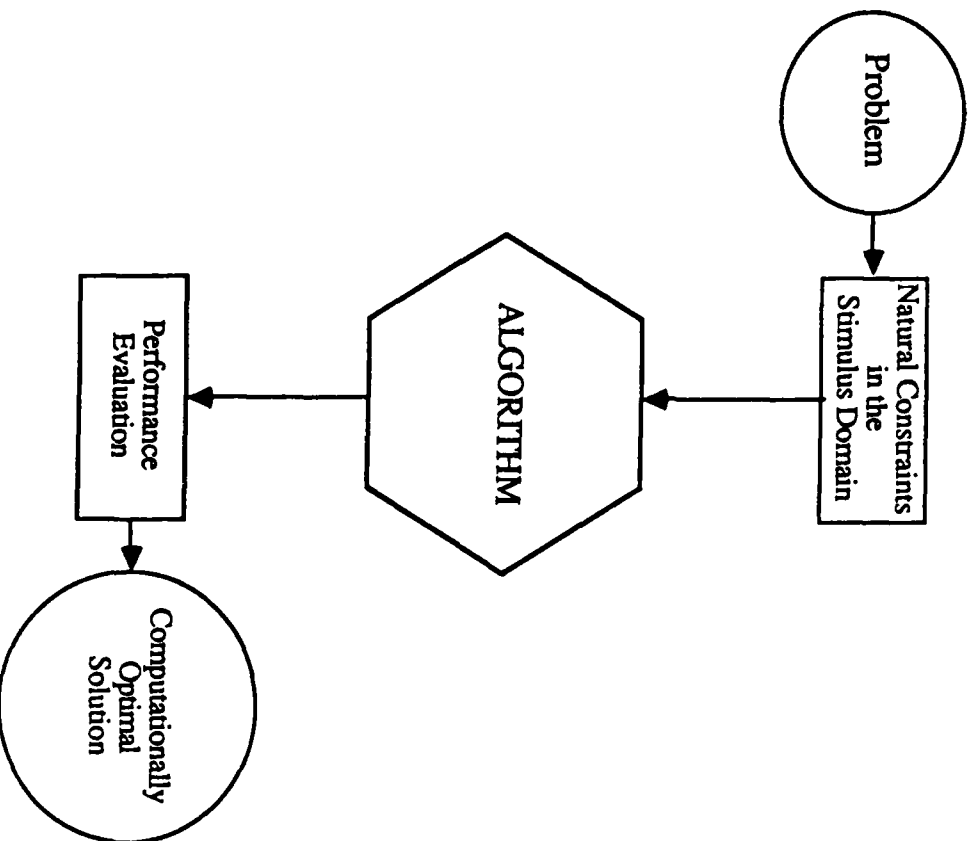
Understanding {
SA = Scene Analysis
IU = Image Understanding
CV = Computer Vision

A HISTORY OF COMPUTER VISION

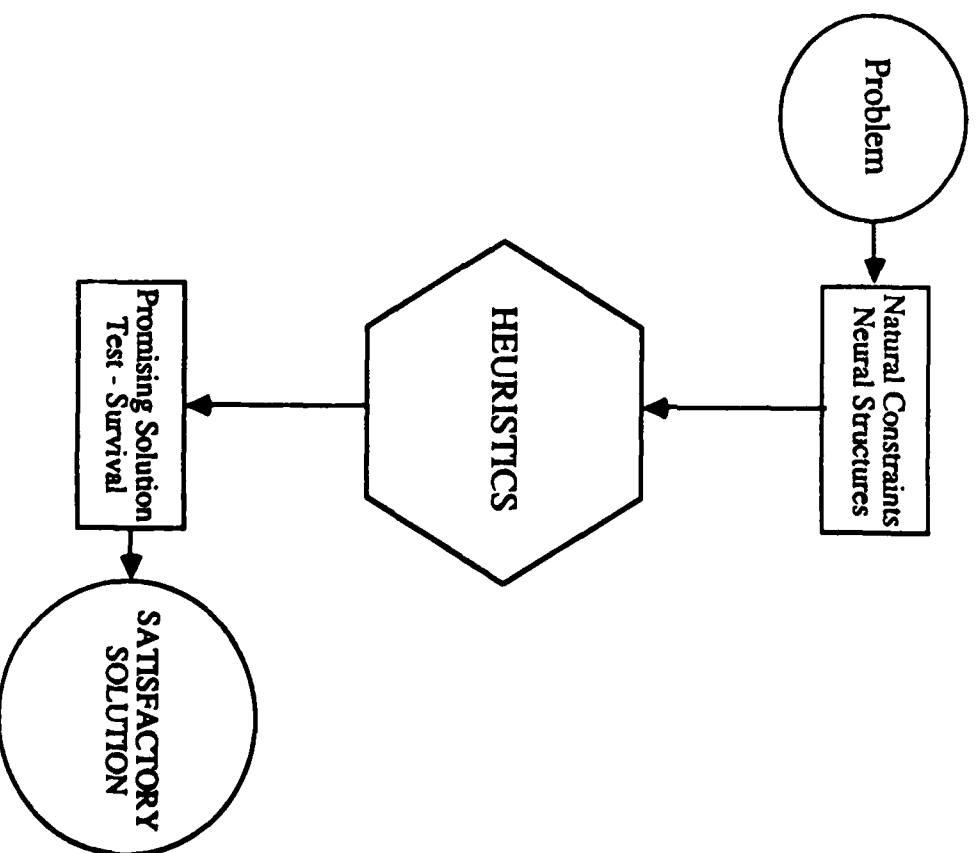


APPROACHES TO VISION

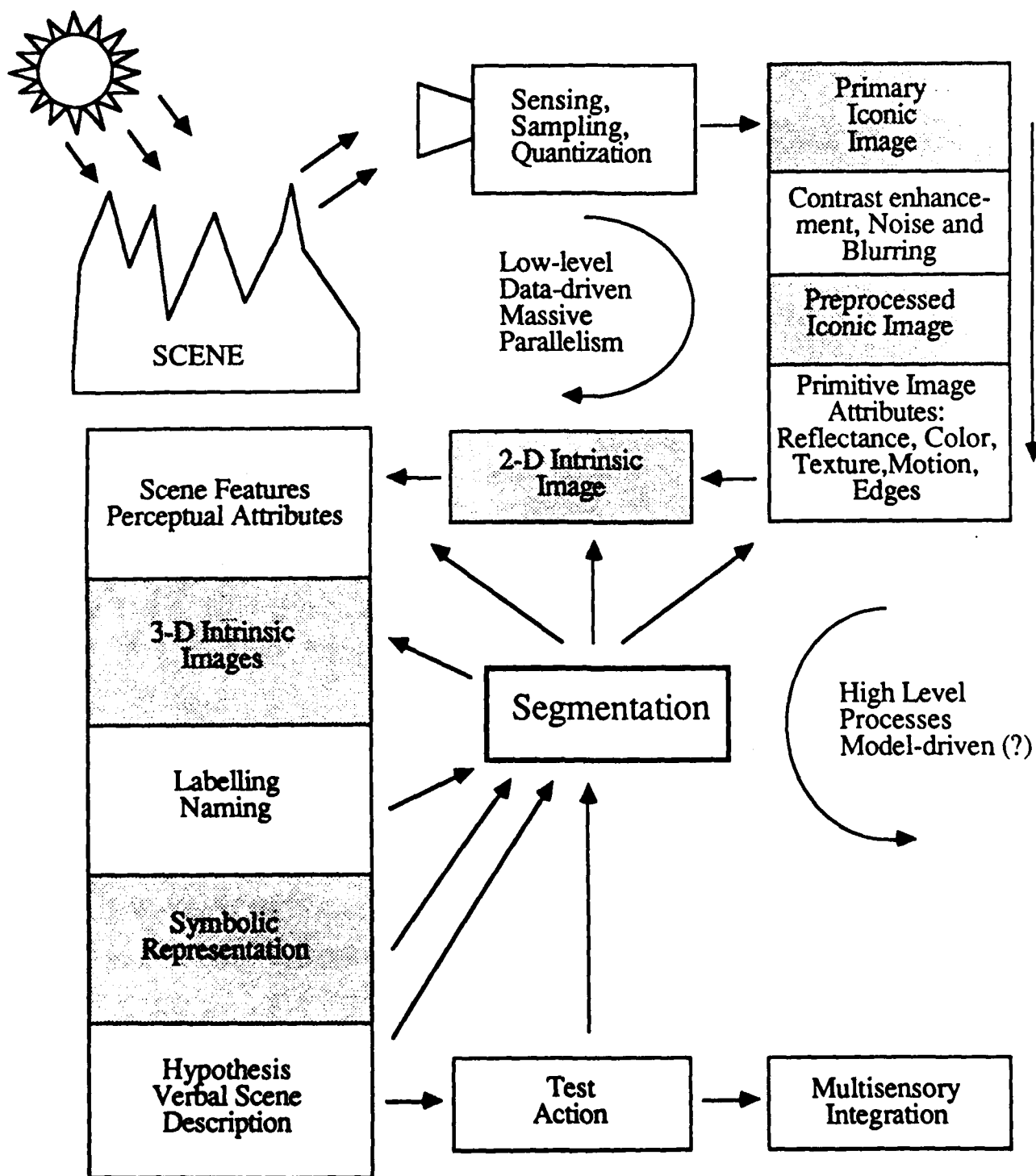
Vision through Synthesis



Vision through Analysis

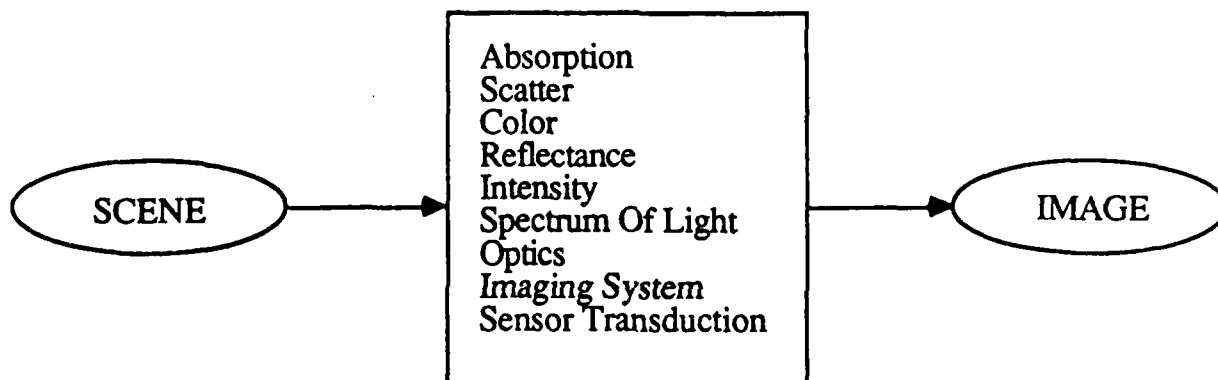


REPRESENTATIONAL SCHEME OF MACHINE VISION



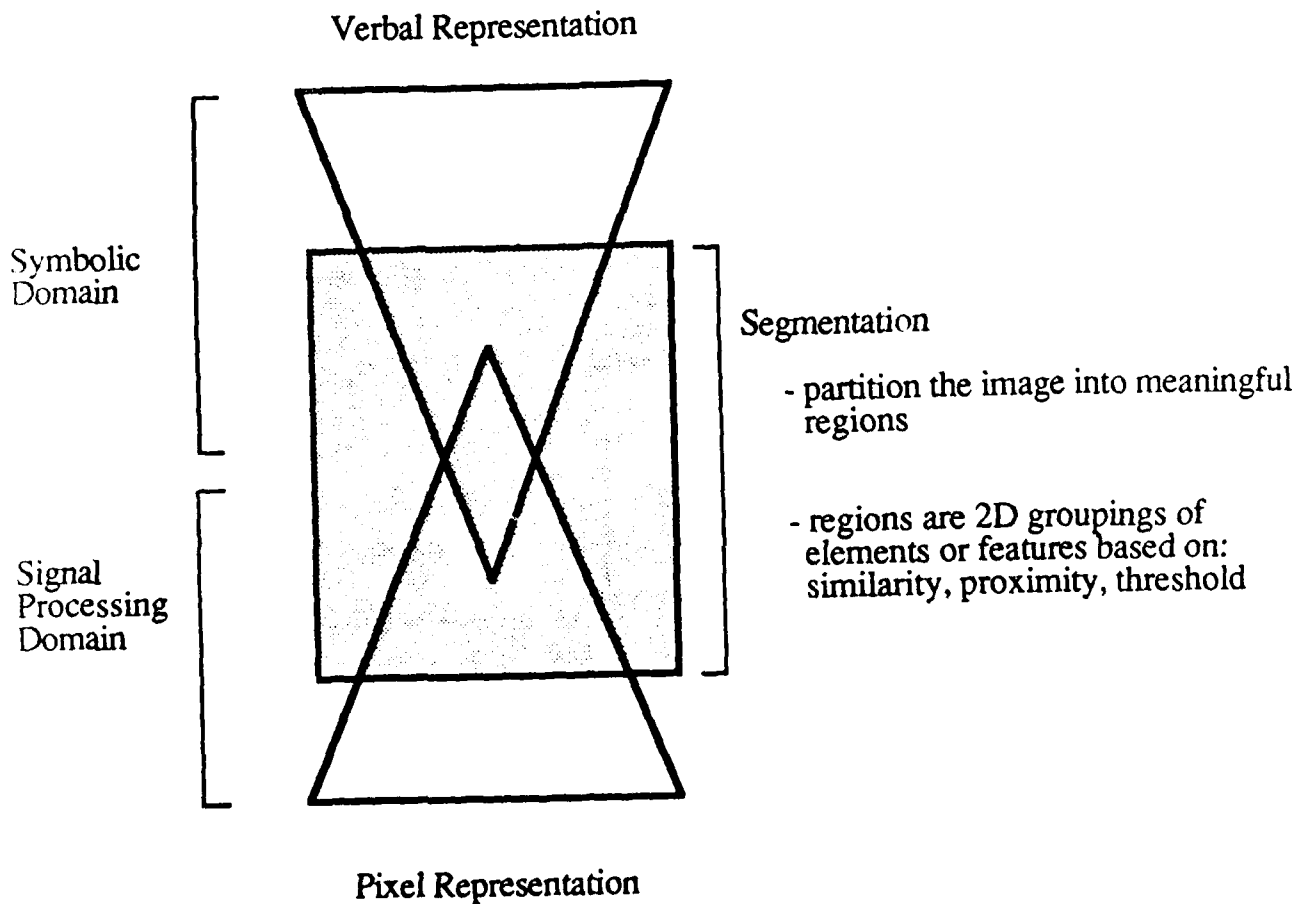
COMPUTER VISION - THE BASIC PROBLEM

- Information is lost during image formation
- Inverse mapping from images to scenes is not unique



- Image understanding is difficult because we do not understand how the various confounding processes interact in the resulting image
- Image understanding is difficult because we do not understand understanding

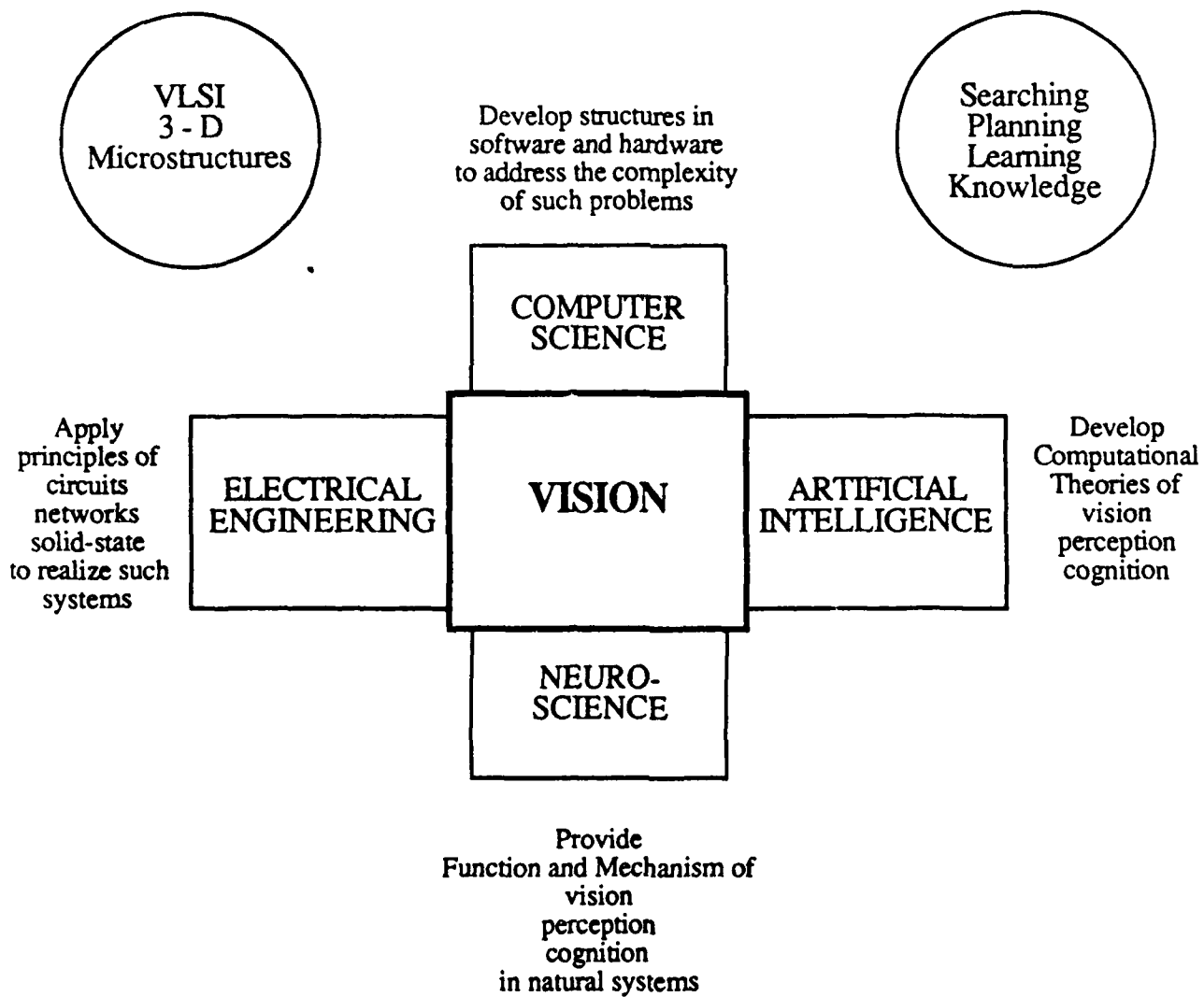
THE SEGMENTATION PROBLEM



Segmentation is difficult

- Conceptually
 - Data driven?
 - Knowledge driven?
 - Both?
- Computationally
 - Transition in the data structure from intensity to symbols

VISION MUST BE INTERDISCIPLINARY



COMPUTATIONAL NEUROSCIENCE

ASSUMPTIONS

BRAIN = information processsing system

*known properties of neurons can be simulated
with available technology*

*information processing capabilities of neural system
do not involve super-natural principles (monism - dualism)*

REDUCTIONIST VIEW

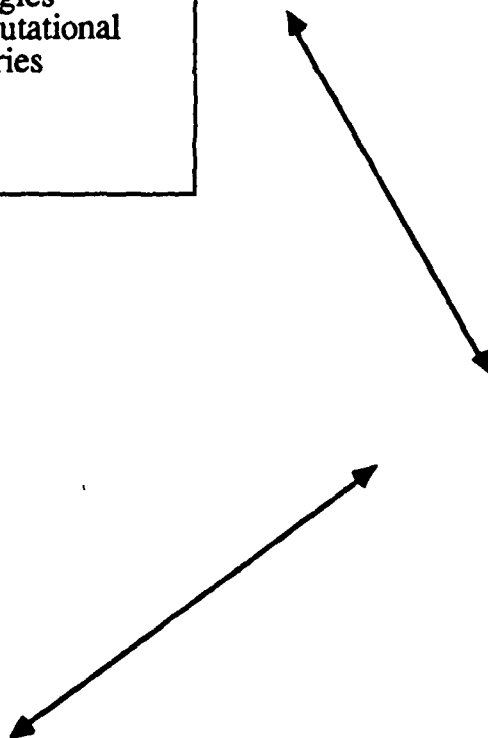
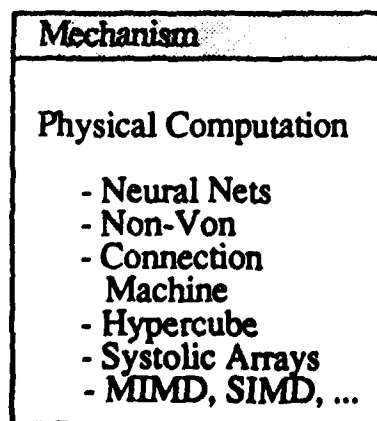
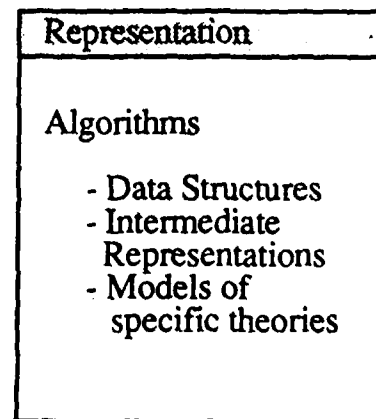
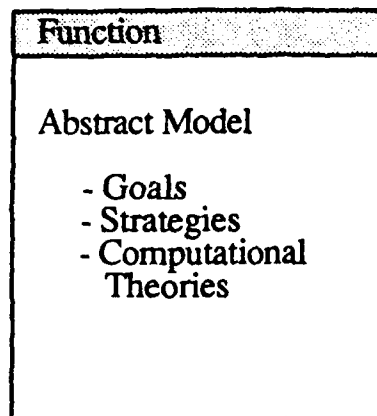
Essential features of brain functions are based on:

CONNECTIVITY

SIGNAL PROCESSING

NETWORK PLASTICITY

COMPUTATIONAL NEUROSCIENCE VIEW OF IMAGE UNDERSTANDING



REAL TIME REQUIREMENTS

Perception has real time constraints

Current realizations of machine perception are slow

Need New Approach -

dedicated parallel architectures and distributed computation

Implication -

INSEPARABILITY OF STRUCTURE AND FUNCTION

WHAT ABOUT PREVIOUS WORK?

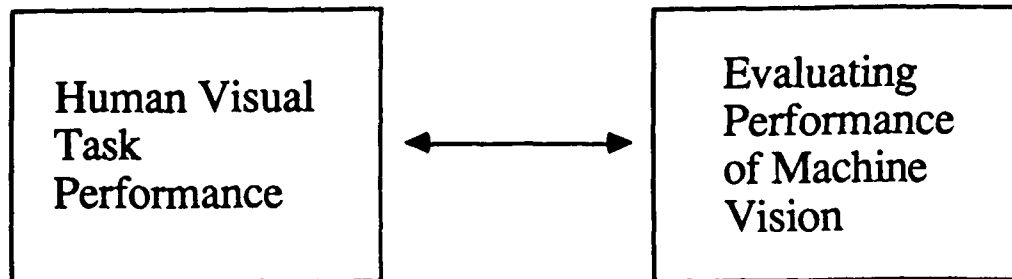
- DO THEY REALLY WORK?
- HOW WELL?
- IN WHAT DOMAINS?

DON'T KNOW -

Because to evaluate machine vision systems is
to know what vision is.

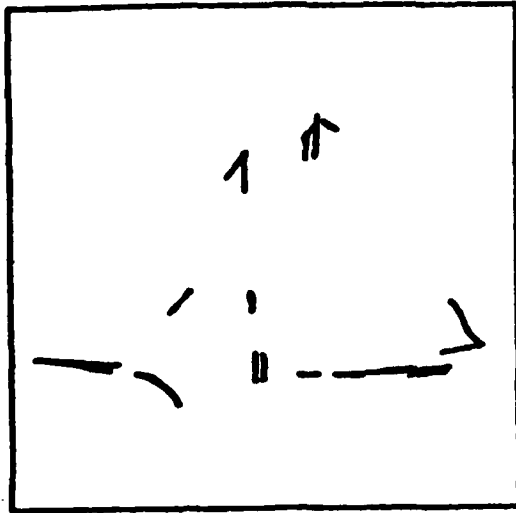
CATCH - 22

OUR APPROACH



Need tasks that meaningfully compare human
and machine vision

GESTALT CLOSURE TEST

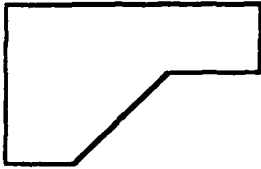


- Segmentation task
- Incomplete data to match against a model

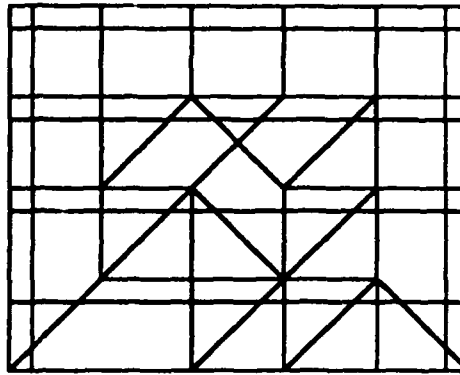
What is the model?

How much data is needed?

HIDDEN FIGURES TEST



FIND
THIS
FIGURE

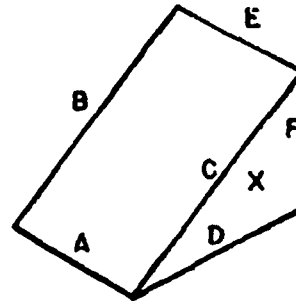
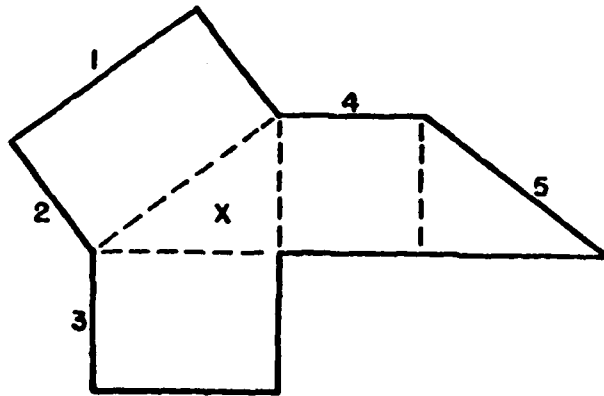


IN THIS DRAWING

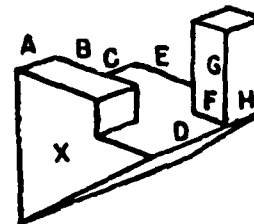
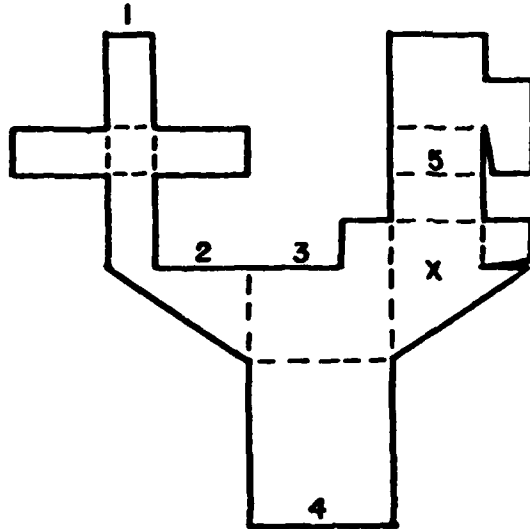
- Hard for people, easy for machines
- Not an appropriate visual task for comparison
comparison and evaluation

SURFACE TESTS

Some are easy:



Some are hard:



- Rotation and manipulation of internal models
- Easier ones can be solved by labelling techniques
- Harder ones require ?

CONCLUSION

- New approaches needed for machine vision
- How do you evaluate current and future systems?
- Develop evaluation model based on human visual performance