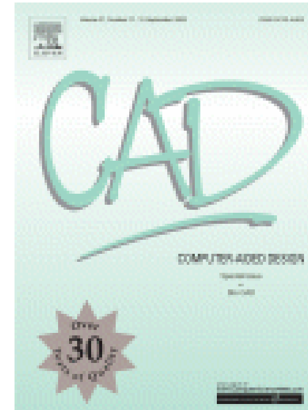
3+ Decades of Searching CAD Data: Historical Perspective and Future Directions

William C. Regli
Department of Computer Science
Department of Mechanical Engineering & Mechanics
Department of Electrical & Computer Engineering
College of Engineering
Drexel University
3141 Chestnut Street
Philadelphia, PA 19104



CALL FOR PAPERS

Special Issue of Computer-Aided Design

Shape Similarity Detection and Search for CAD/CAE Applications

Guest Editors: William C. Regli, Drexel University, USA
Michela Spagnuolo, IMATI-CNR, Genova, Italy

1 October 2005	**Deadline for Submission of Full Papers**
1 February 2006	**Deadline for Submission of Revised Papers**
Summer 2006	**Expected Publication**

3D CAD data has become the currency of the modern engineering enterprise. As the industry has matured, CAD models have become ubiquitous artifacts that provide high-fidelity descriptions of engineered objects as well as capture

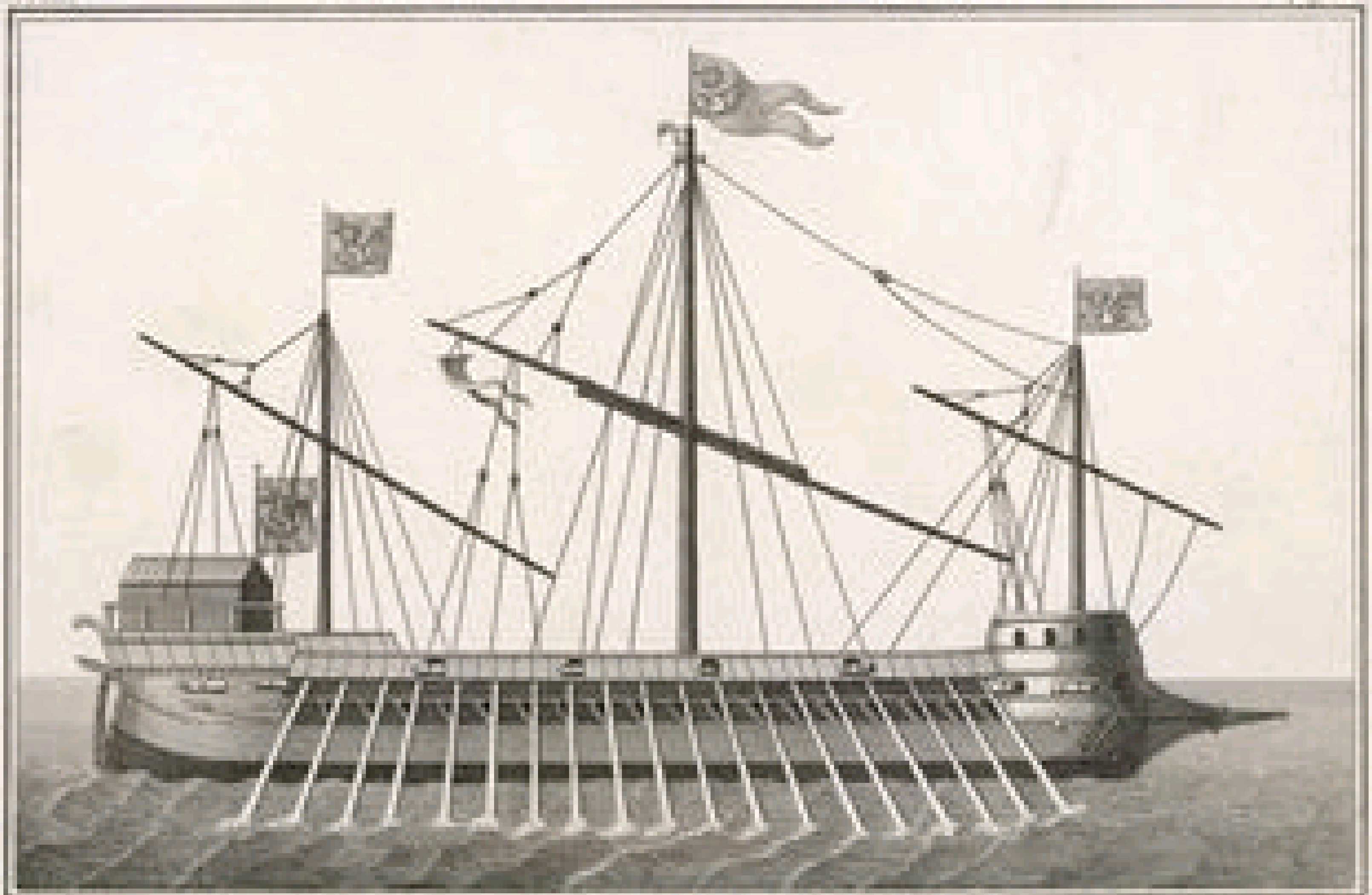
- Integration of CAD search with database and PLM systems
- Applications, prototypes, and fielded systems
- Empirical studies, scalability results and benchmarks

For this special issue, the Computer-Aided Design and the guest editors strongly encourage potential authors to use models and datasets that are publicly available. Additionally, authors may provide online links to models they use in their papers and experiments. Results presented must be independently verifiable or reproducible. Datasets appropriate to this issue include those noted in the CAD "Information for Authors" such as the Design Repository (http://www.designrepository.org), the Princeton Shape

Goal of the Talk:
Generate Discussion

- What are the challenges for modern engineering record keeping & data management?
- What is the current state of the art in engineering record keeping & data management?
- What is the role of 3D search in engineering data management?
- What are the emerging challenges and limitations of current technologies?

A Venetian Galleas at the battle of Lepanto.

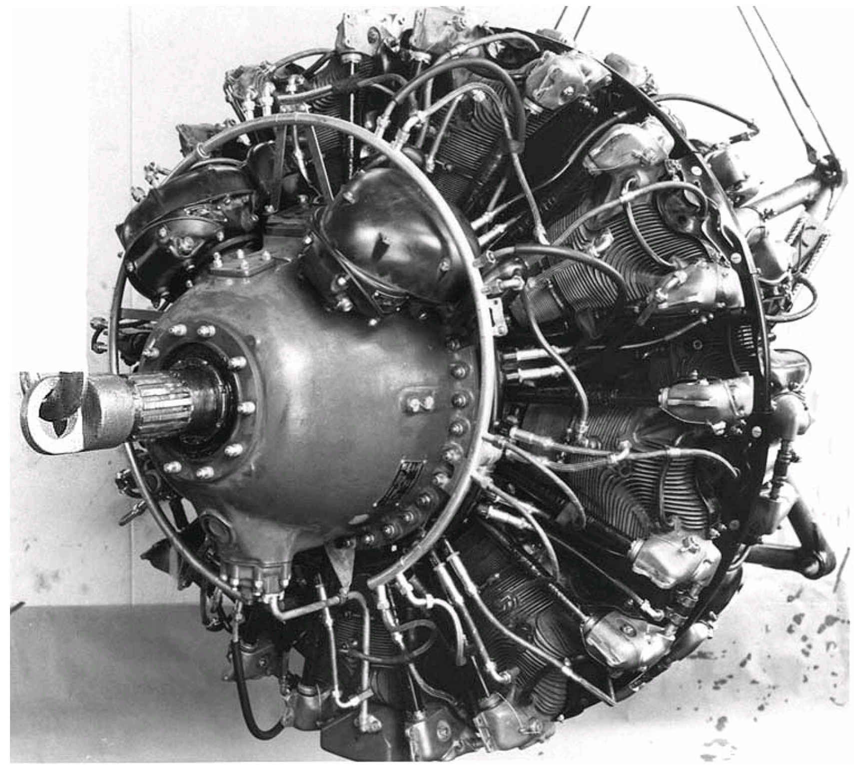


Arsenale di Venezia

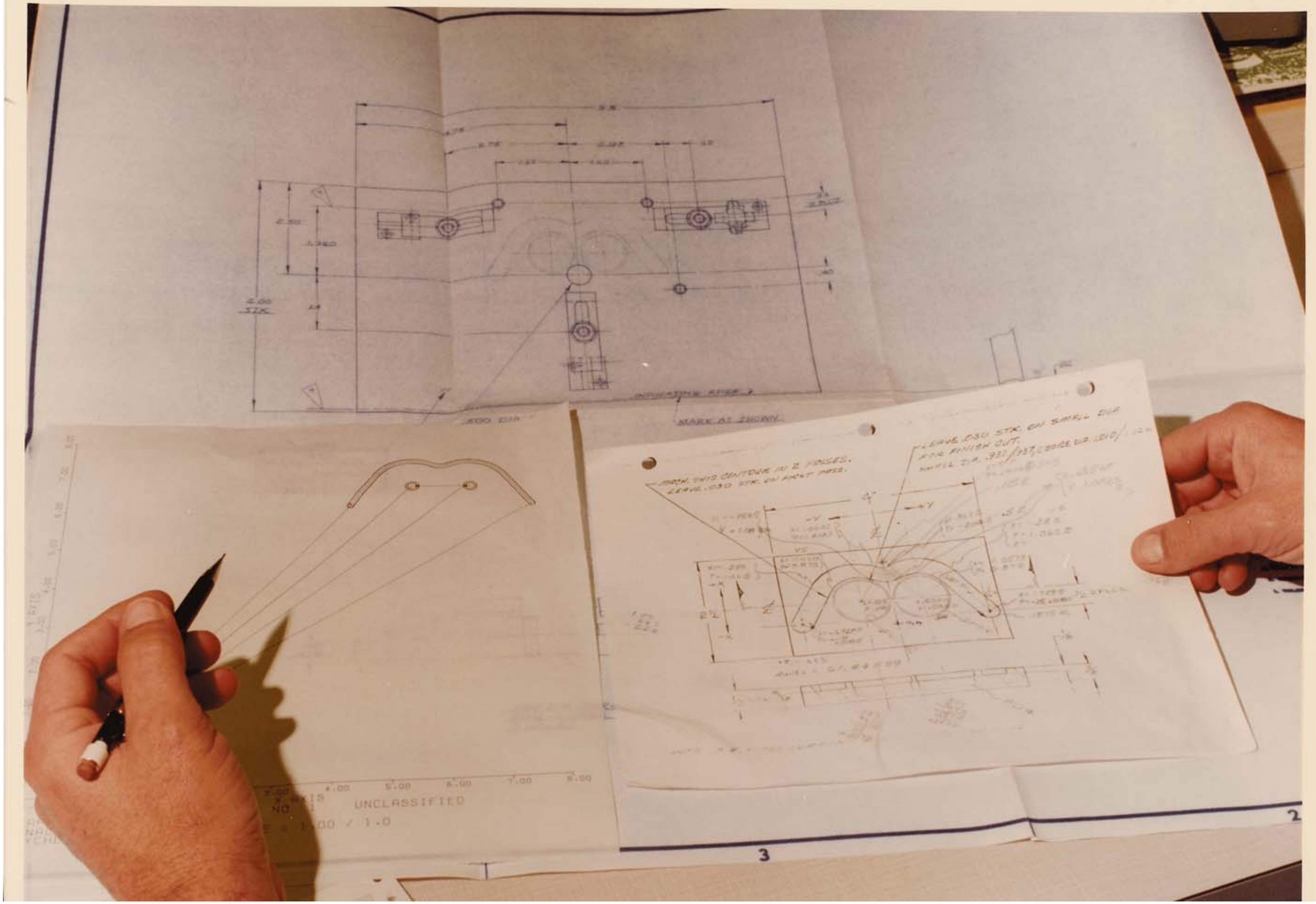
- Name “Arsenale” from the arabic *Dar al Sina’a*
- By 1104 it had:
  - An assembly line
  - Mass production facilities
  - Standardized parts
  - Developed frame-first ship construction methods
  - Repair and refit techniques
- Employed ~16,000 people
- Could build a ship from scratch in less than 1 day

Fast Forward 800 years… Pratt & Whitney R-2800 Double Wasp

- An 18-cylinder two-row radial engine providing up to 2,500 horsepower. An important factor of Allied air supremacy in World War II, the Double Wasp remained in production until 1960.
  - Displacement: 2,804 cubic inches
  - Revolutions per minute: 2,250-2,800
  - Weight: 2,360 pounds
  - First run: 1937
  - First flight: 1939
  - Production years: 1939-1960
  - Engines produced: 125,334
- Over 10,000 parts

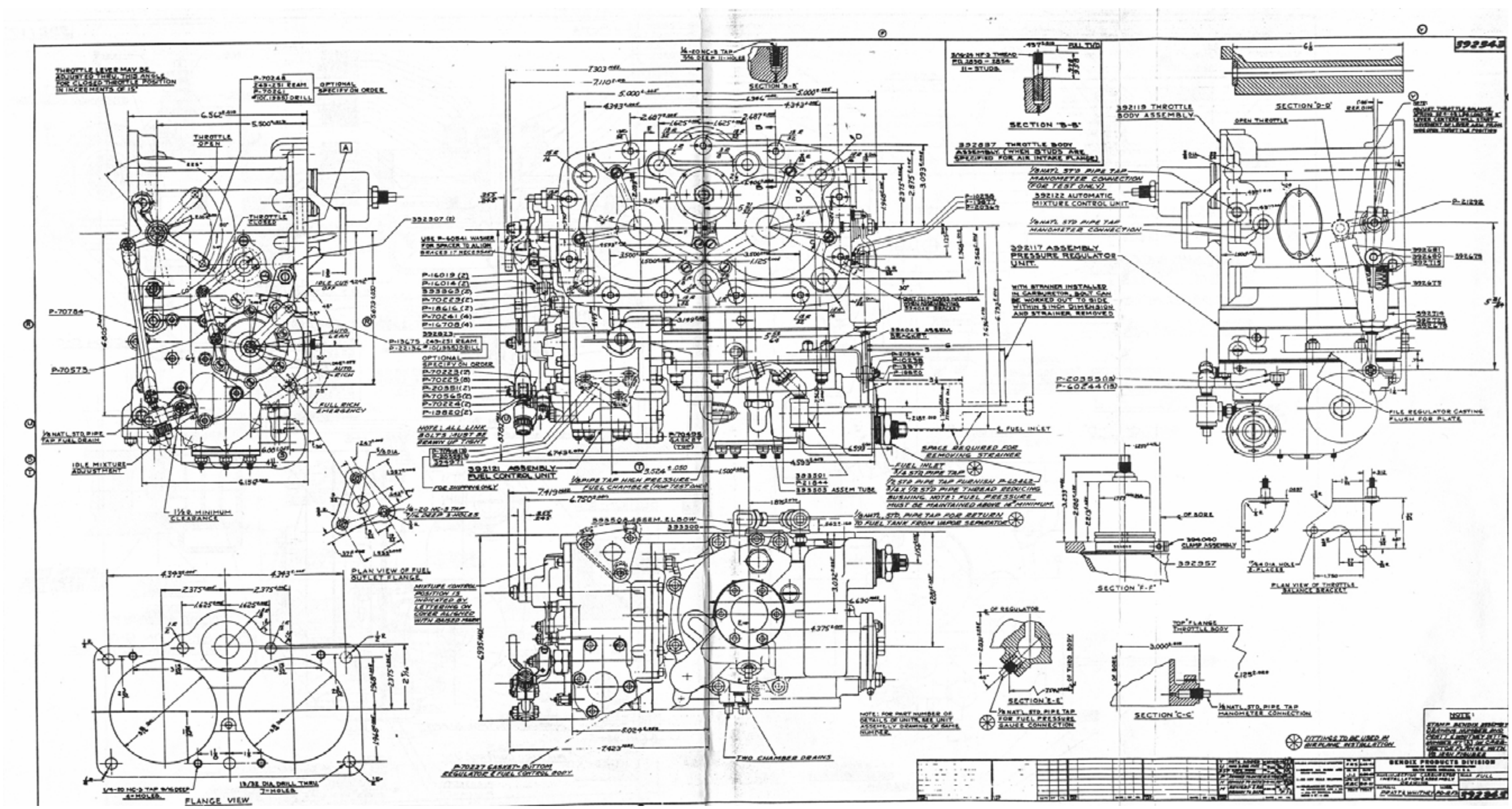


The Kansas City Plant



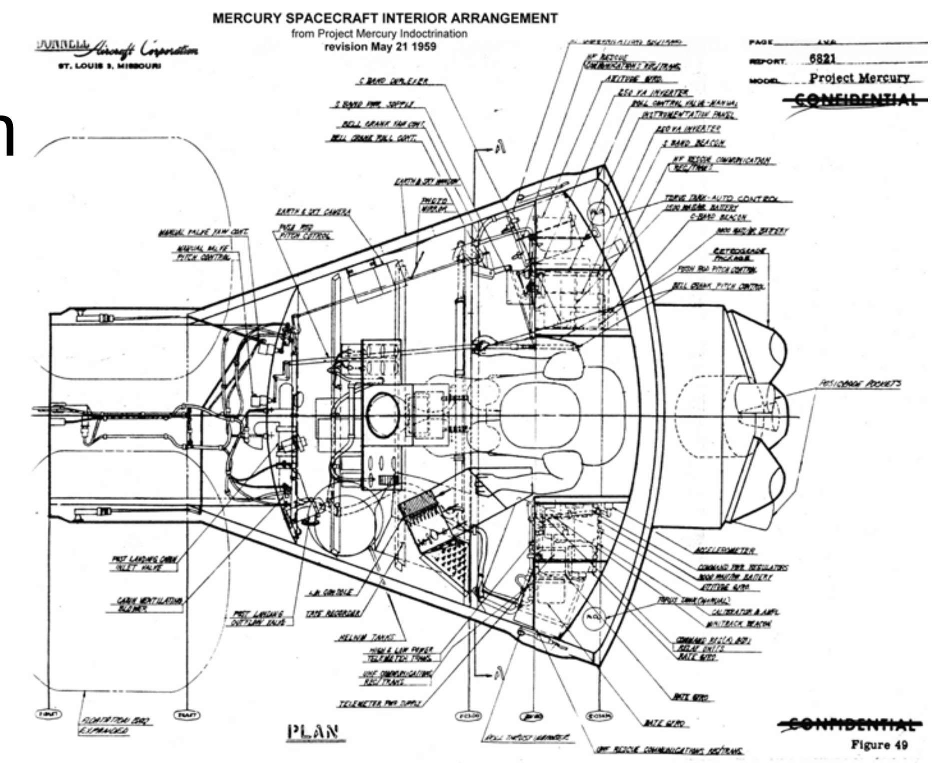
UNCLASSIFIED
X AXIS
MARK AS SHOWN
LEAVE .030 STK. ON SMALL DIA FOR FINISH CUT.
3
2

Traditional Design Record Keeping as Art Appreciation



Traditional (Current) Design Record Keeping

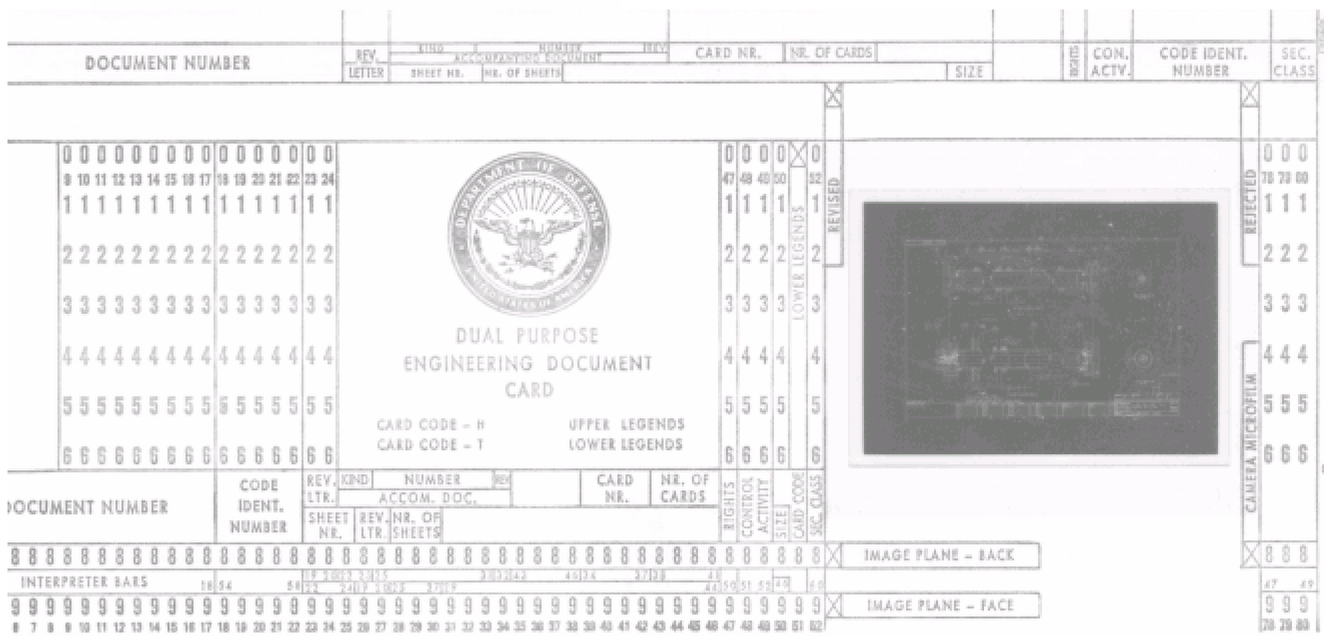
- Paper-based workflow process
- For Human-to-Human exchange
- CAD is *fancy drafting*
- CAD does not capture design semantics (beyond geometry)



Designed in 1959

CAD Databases

- Aperture cards
- Physical files



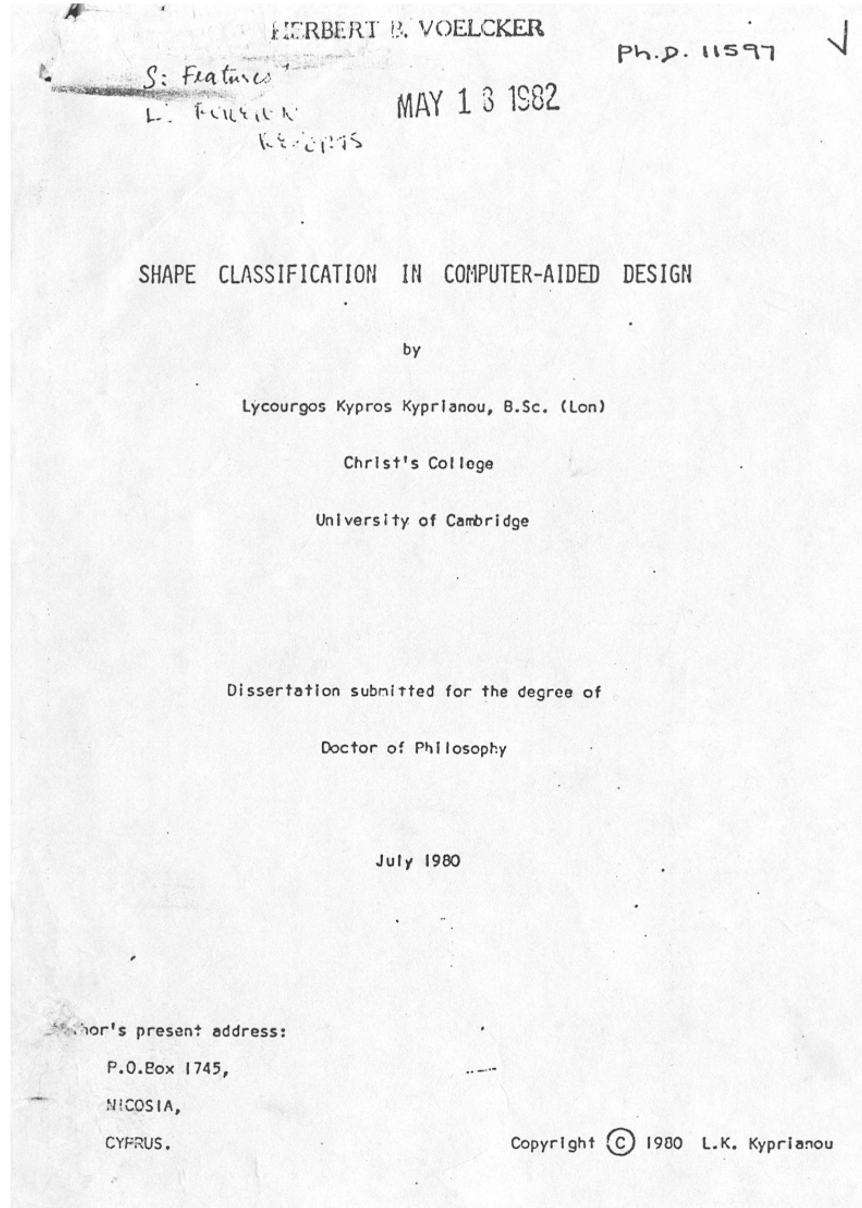
Modern History

- 1963
  Ivan Sutherland's
  Sketchpad
- Modified oscilloscope for drawing
- The original CAD system
- Sidenote: His advisor was Claude Shannon

Courtesy Marc Levoy @ Stanford U

CAD Search

- Some of the original work on Features was for Shape Classification
  - Kyprianou 1980
- Goal: Creation and Search of CAD Databases



HERBERT B. VOELCKER

Ph.D. 11597

S: Features

MAY 13 1982

SHAPE CLASSIFICATION IN COMPUTER-AIDED DESIGN

by

Lycourgos Kypros Kyprianou, B.Sc. (Lon)

Christ's College

University of Cambridge

Dissertation submitted for the degree of

Doctor of Philosophy

July 1980

...hor's present address:

P.O.Box 1745,

NICOSIA,

CYPRUS.

Kyprianou's Classifier

- Based on a set of morphological features
- Used a grammar to parse features from solid model brep

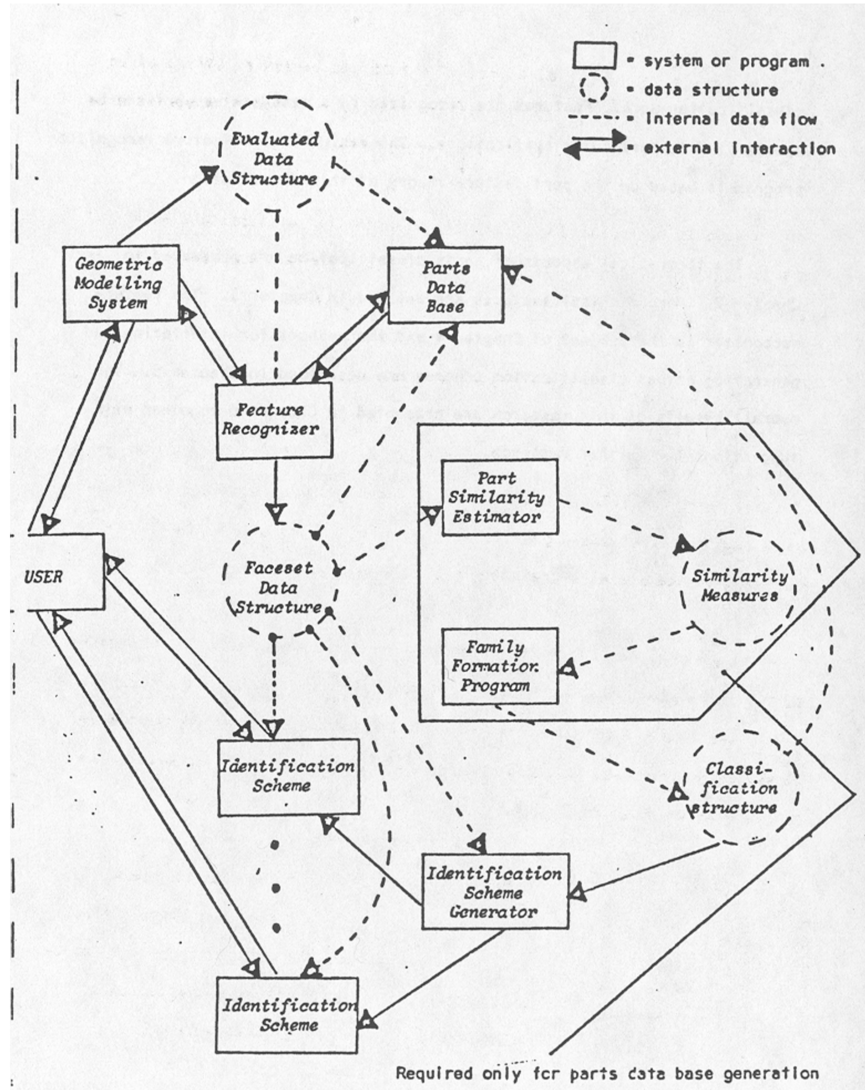
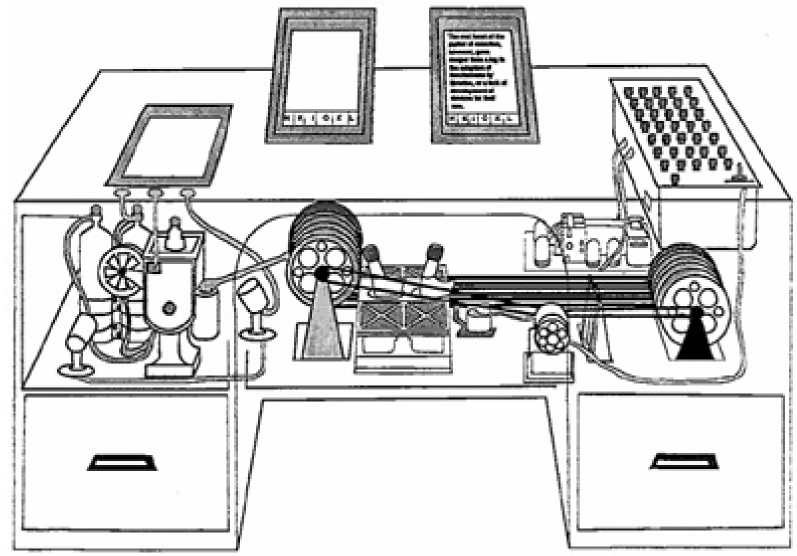


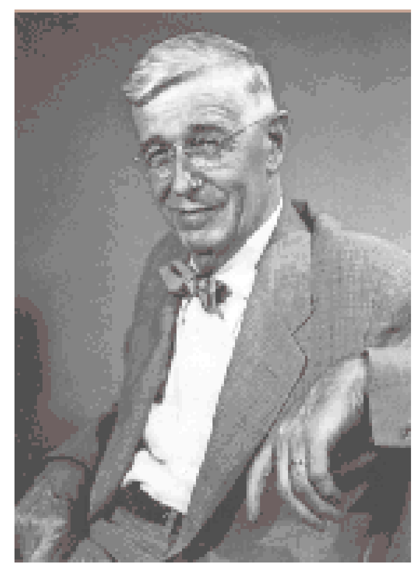
Fig. 1.11 An Automatic Parts Classification System

- 23 -

Sutherland… influenced by V. Bush

- *As we may think,*
  Atlantic Monthly, July 1945
- Proposed fantastic device: The Memex
- Article predicted
  - Digital photography & storage
  - Hyertext
  - Speech recognition
  - Personal computers, Internet
- Influenced many, still today
  - E.g. Google





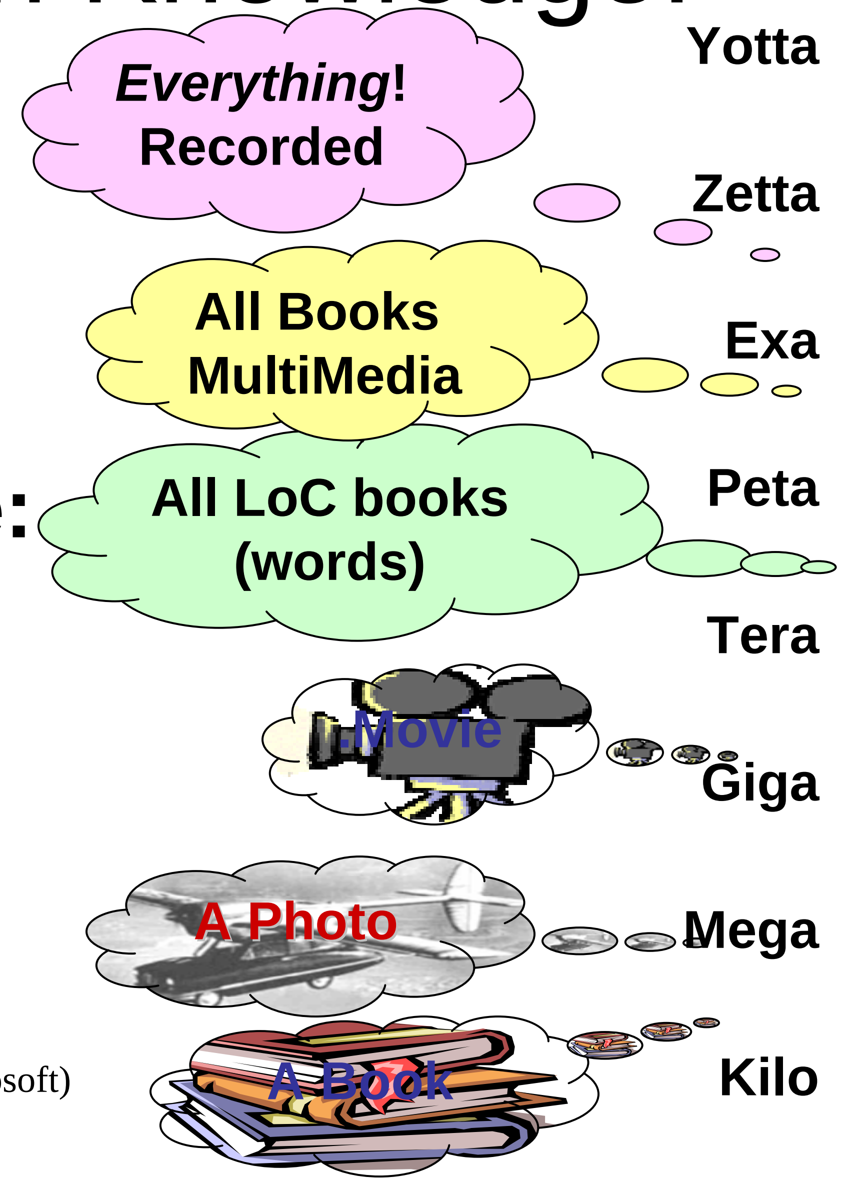
Vision of Memex: Digitization of the Totality of Human Knowledge!

- **Soon everything can be recorded and indexed**

- **Most data never be seen by humans**

- **The Precious Resource:**
  Human Attention!
  Auto-Summarization
  Auto-Search
  is key technology.

www.lesk.com/mlesk/ksg97/ksg.html



24 Yecto, 21 zepto, 18 atto, 15 femto, 12 pico, 9 nano, 6 micro, 3 milli

(From the Turning Award Lecture of Dr. James Gray, Microsoft)

Future Design Repositories

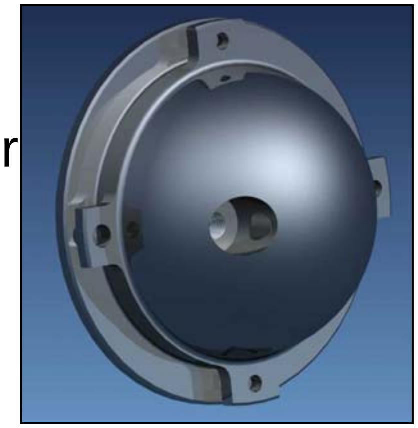
- Design Knowledge
  - capture, index and archive, reuse
  - models, simulations, analysis, revisions, maintenance and performance, alternatives and dead-ends, process and workflow, rationale and history
  - capturing the "***Why?***"

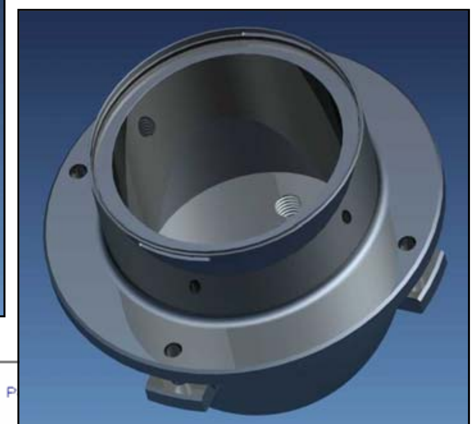
Designed in 1949-52

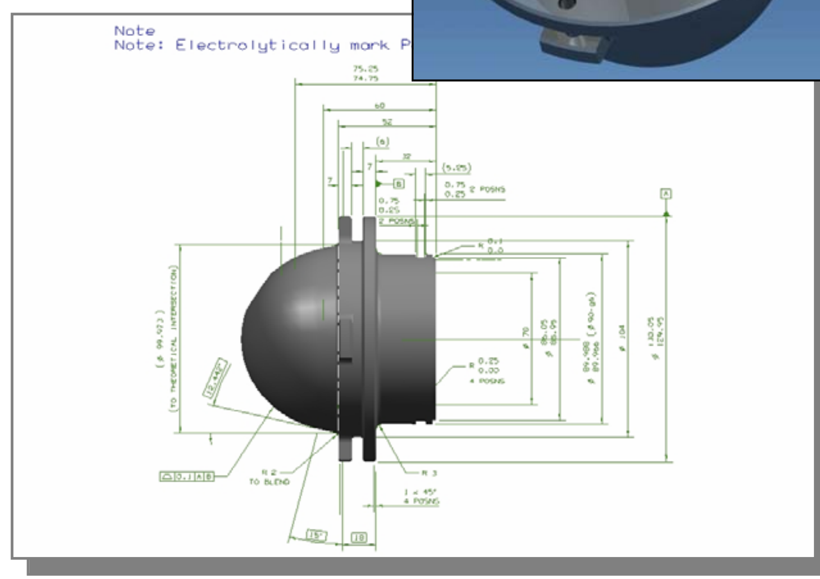
Expected Service Lifespan:
1954-2045
91 Years!

Consider a contemporary example: **<u>UK AWE Amber 2</u>** Part

- Partner: **Kansas City Plant**
  - 50 year history
  - Primary manufacturing facility for the DoE & NNSA
  - Expertise in discrete parts, electronics, MEMS, …
- The <u>Amber 2</u> part
  - High-precision machined part
  - Designed in the UK
  - Analysis at both UK AWE and KCP
  - Fabricated at KCP



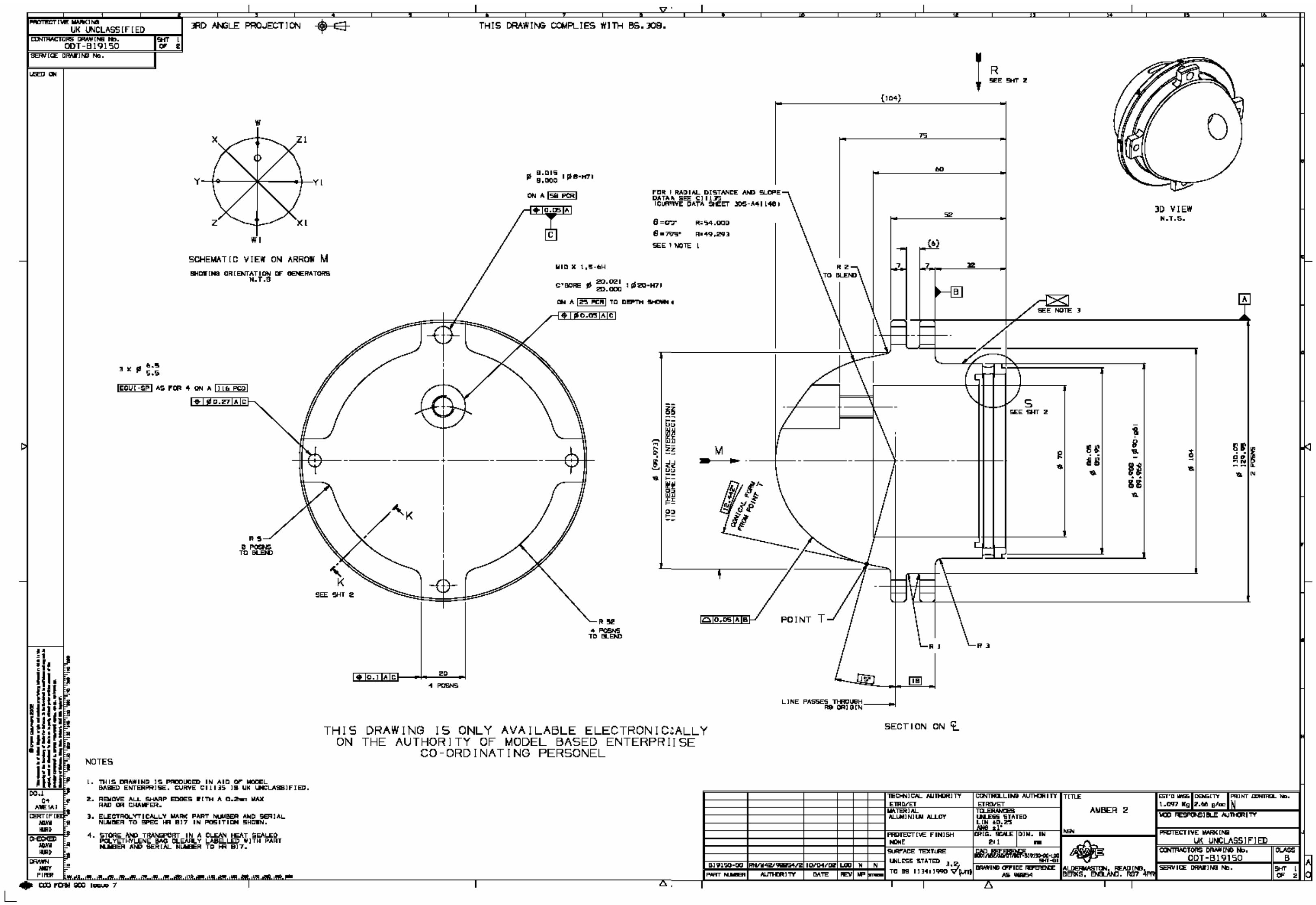




UK AWE Amber 2 Part Data

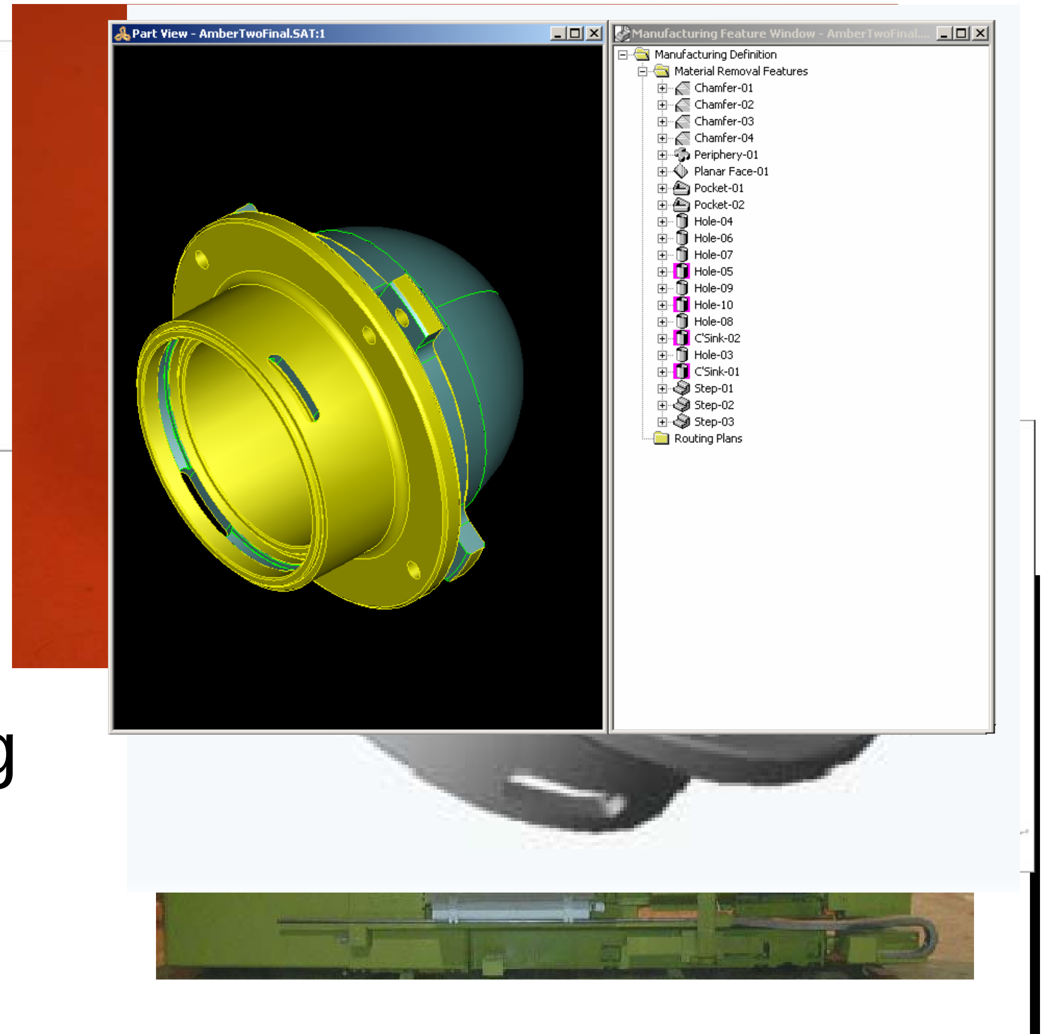
- 2D CAD Drawings
  - TIFF images
- 3D CAD data
  - Parasolid, Pro/E, STEP, ACIS, …
- Shape data
  - Mesh & point cloud
- Tolerance data
  - ASME Y14.5 tolerances and tolerance features
  - Tolerance analysis
- Analysis data
  - FEA parameters and output
- Manufacturing data
  - Features
  - Process plans
  - Manufacturing plan simulations
- Fabrication data
  - Tooling, cost, time
- Inspection data
  - Inspection plan, robotic simulation
- Documentation
  - MS Word files
  - AVIs, MPGs
  - Other files

Current Format of Record



Missing Information

- Tolerances
- Manufacturing Planning
- Analysis
- Inspection
- Fabrication
  - Okuma LH35-N CNC lathe
- Reverse Engineering



Limitations of Current Technology

- Technologies such as GT Coding are pre-digital
- Matching gross shape doesn't often help answer meaningful engineering questions
- Important attributes are quasi-geometric and hard to represent
  - Tolerances, material, surface finish, mfg plans etc
- Important features vary by application (and are ill defined)
  - Machining, SFF, cost estimation, etc.
  - Feature Recognition is challenging
- We are still determining the questions that can be answered by CAD search
  - How to query Design Repositories and for what purpose?

Challenges for 3D Search

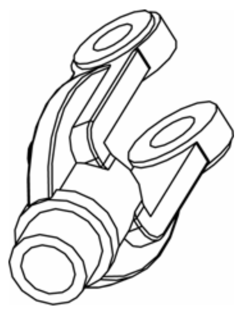
1. CAD objects are different than “shape” objects
2. Focus is on the “query by example” paradigm
3. Techniques limited to individual, discrete, objects
4. Lack of well defined object semantics beyond shape
5. Lack of use cases for applications
6. No accepted procedures for how to measure performance of techniques
   – Most use their own datasets for testing
   – Evaluation procedures are sometimes opaque
   – Metrics for “success” are not standardized

What Makes CAD Objects Different?

- **Different fundamental underlying representation from graphical objects**
- Engineering objects (CAD objects) have a physical realization
- Object “tags” and classifications are not subjective
- Need to support multiple classification schemes depending on the end-user application

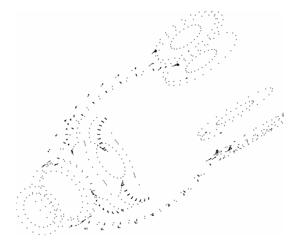
CAD vs Shape Representation

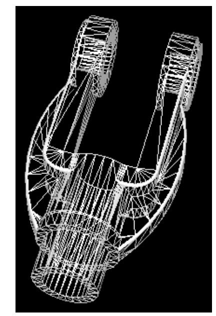
CAD Representation



- Topologically and geometrically consistent
- Implicit and analytic surfaces, NURBS, etc
- Produced using CAD packages and solid modelers

Shape Representation

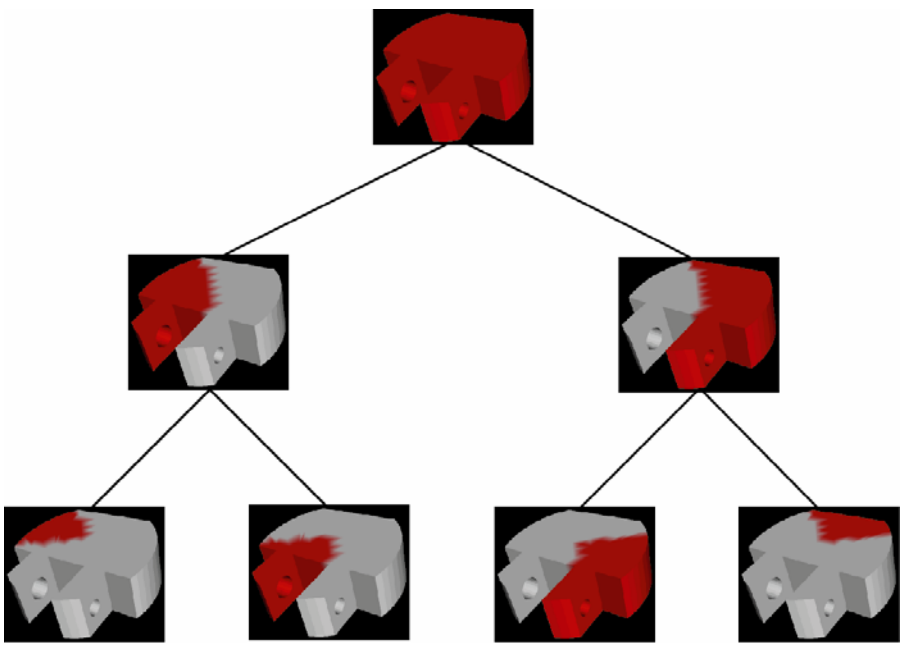




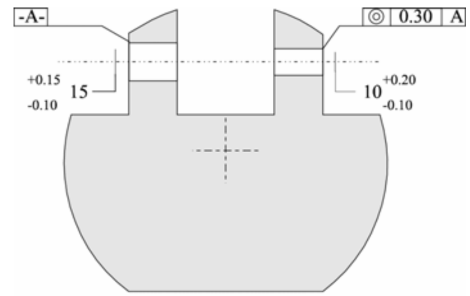
- Approximate representation, error prone
- Mesh or point cloud
- Produced using animation tools, laser scanners

Compare Features

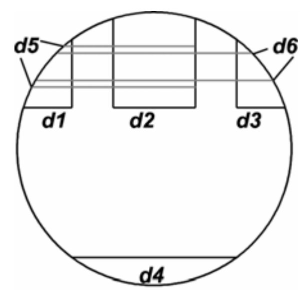
Scale Space

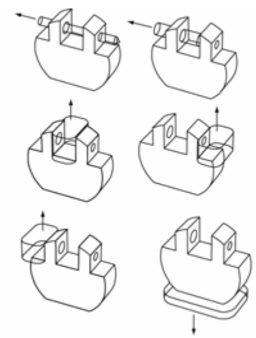


CAD/CAM



t1: datum -A-
t2: dimensional 15 +.15/-.10
t3: dimensional 10 +.20/-.10
t4: co-centricity +/-.30

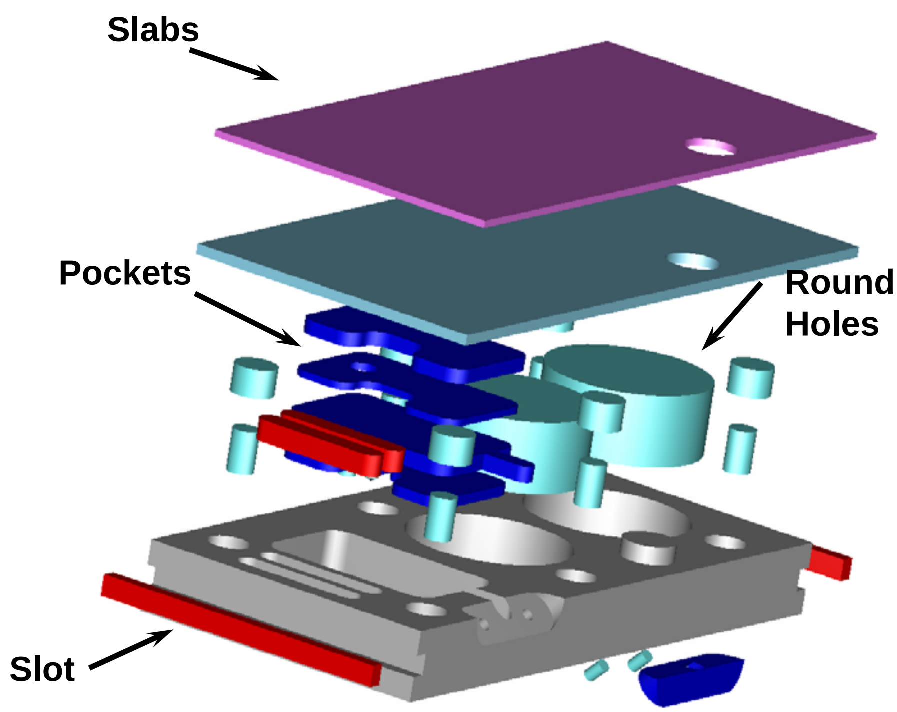




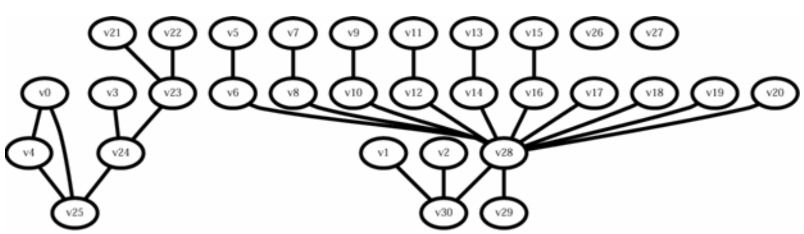
d1: cutout 1
d2: cutout 2
d3: cutout 3
d4: cutout 4
d5: blind hole
d6: thru hole

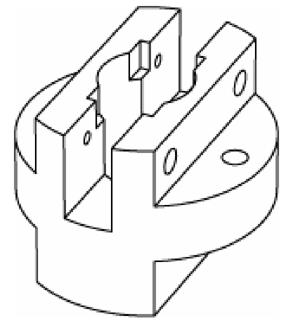
f1: twist drill thru
f2: twist drill, depth 60mm
f3: slot, depth 30mm
f4: slot, depth 20mm
f5: slot, depth 20mm
f6: face mill, depth 10mm

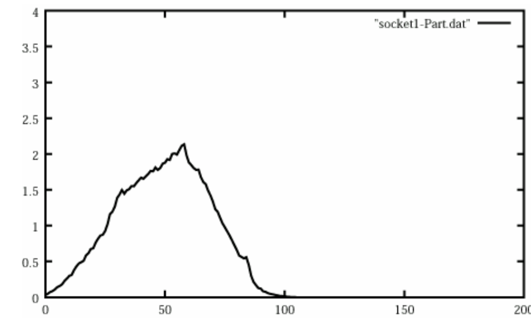
Solid-Based / Shape-Based

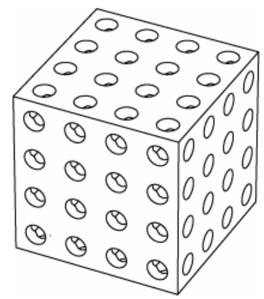


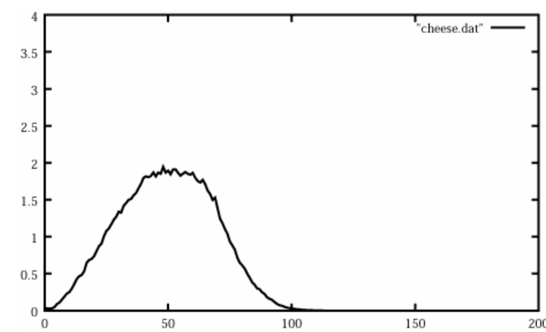
Example: STEP AP 224 Features

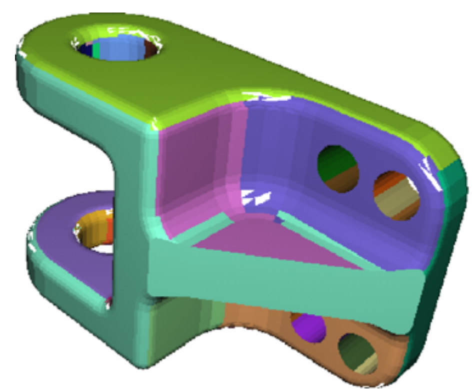










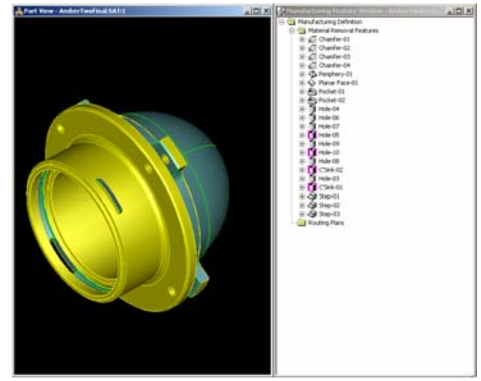


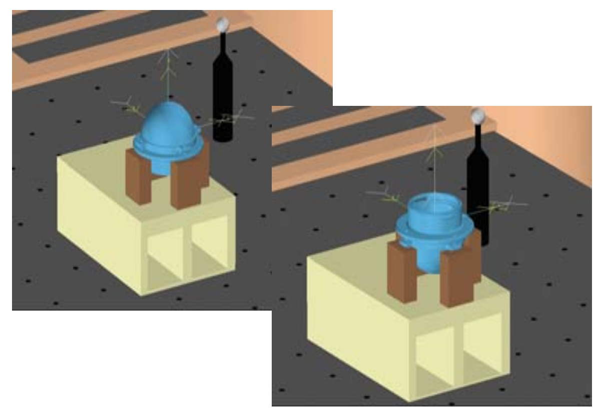
What Makes CAD Objects Different?

- Different fundamental underlying representation from graphical objects
- **Engineering objects (CAD objects) have a physical realization**
- Object "tags" and classifications are not subjective
- Need to support multiple classification schemes depending on the end-user application

Artifacts are Physically Realizable

- Associated with them are
  - Tolerances
  - Manufacturing, Inspection and Analysis Plans
  - Fabrication parameters
    - Okuma LH35-N CNC lathe

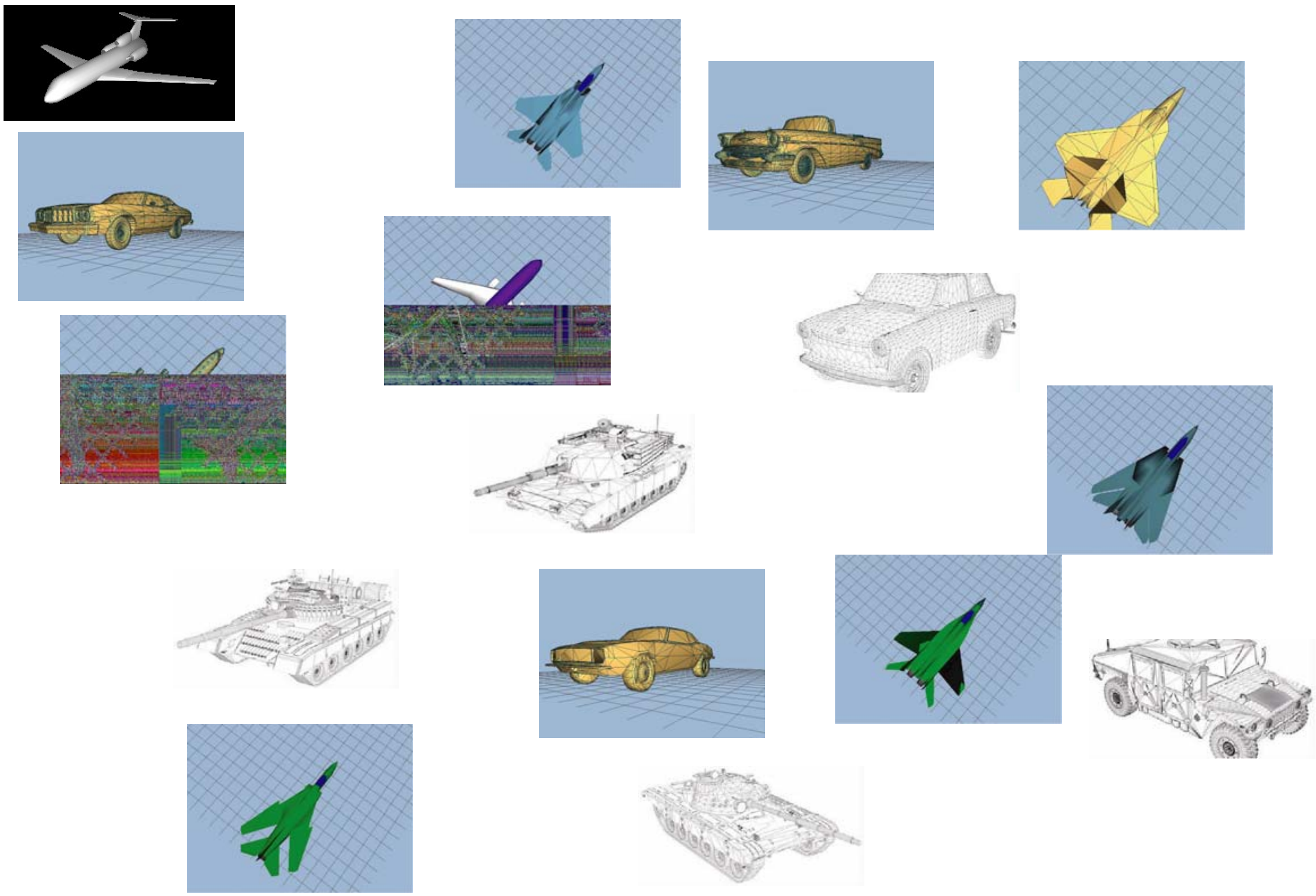




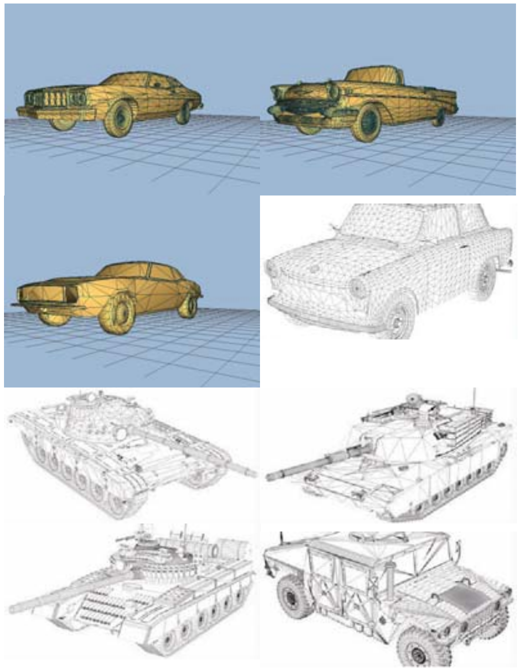
What Makes CAD Objects Different?

- Different fundamental underlying representation from graphical objects
- Engineering objects (CAD objects) have a physical realization
- **Object "tags" and classifications are not subjective**
- Need to support multiple classification schemes depending on the end-user application

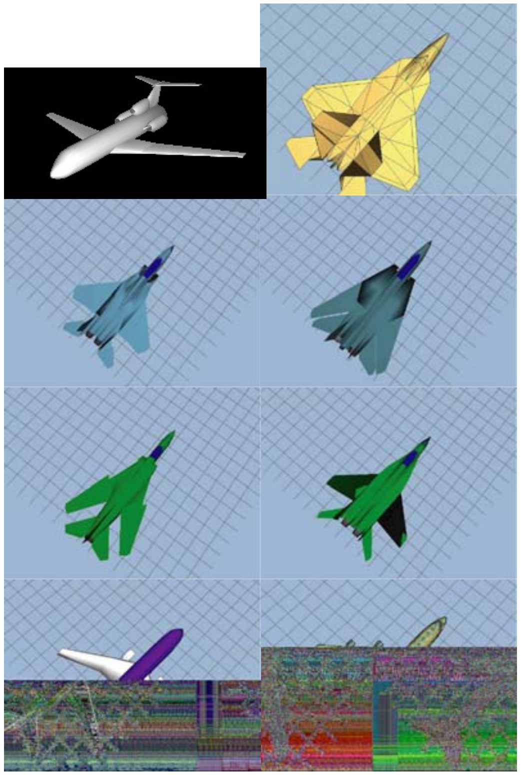
Consider a set of “Vehicles”



Land-Air Classification

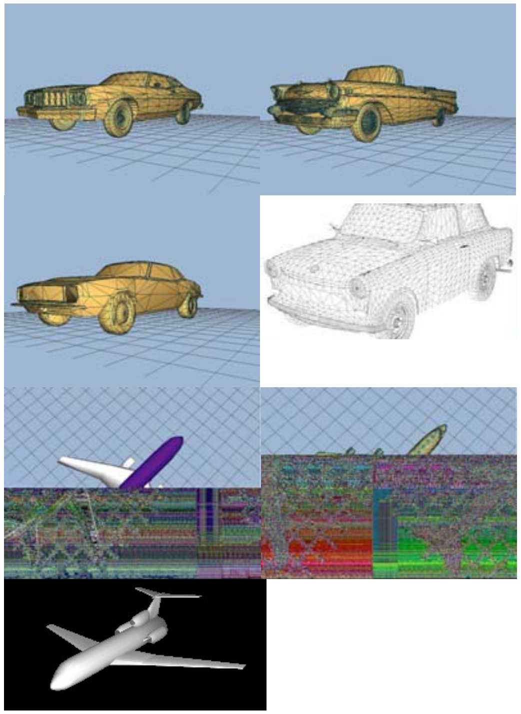


Land



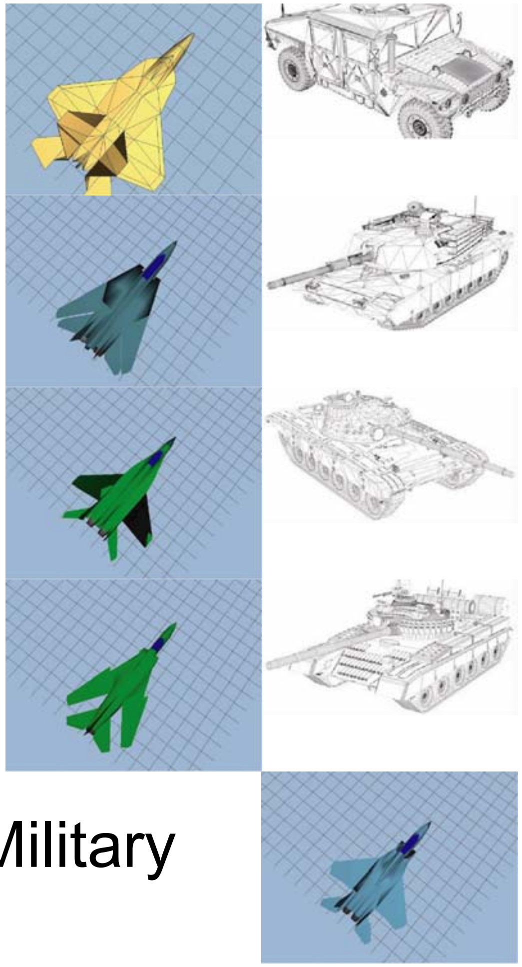
Air

Com-Mil Classification

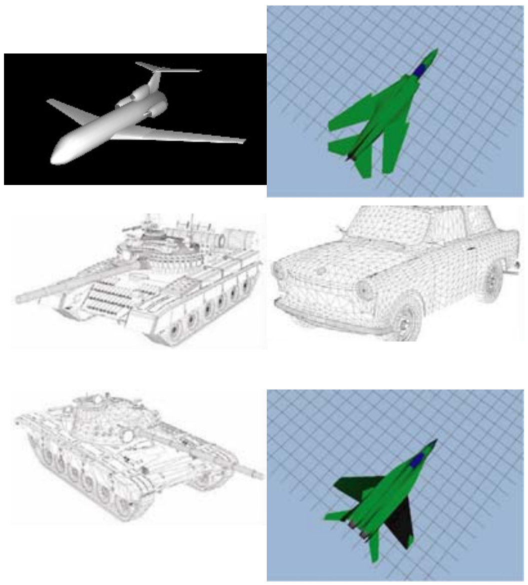


Commercial

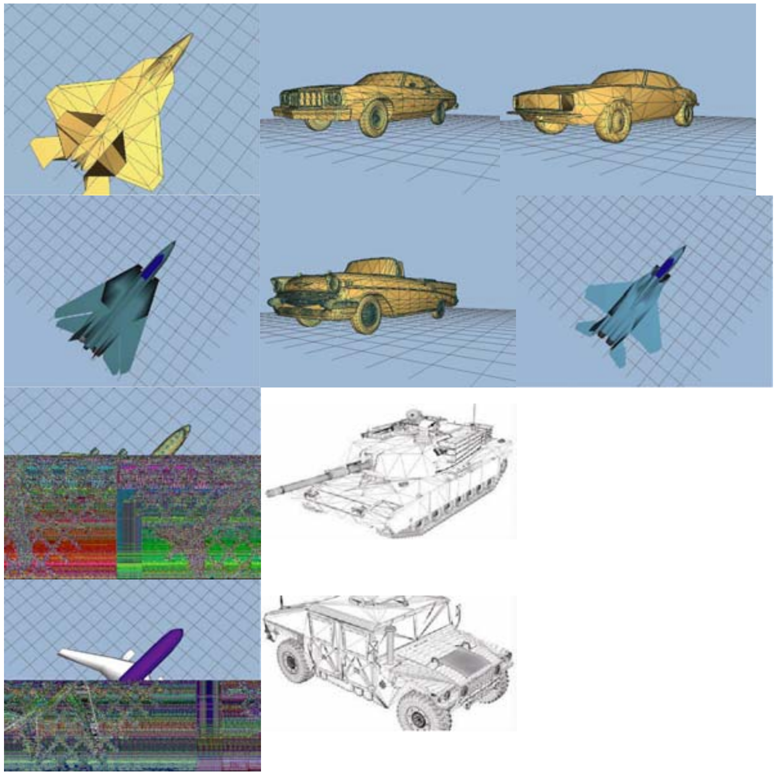
Military



USSR-USA Classification

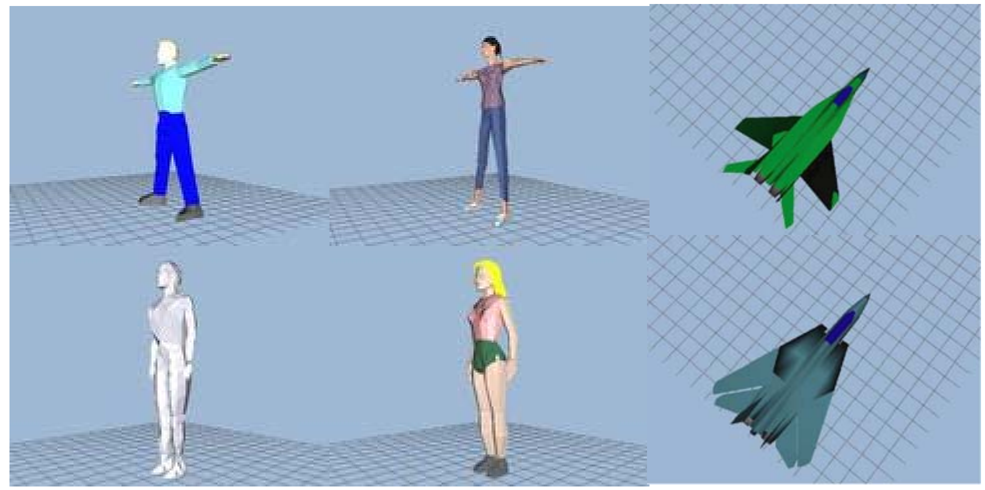


Russian



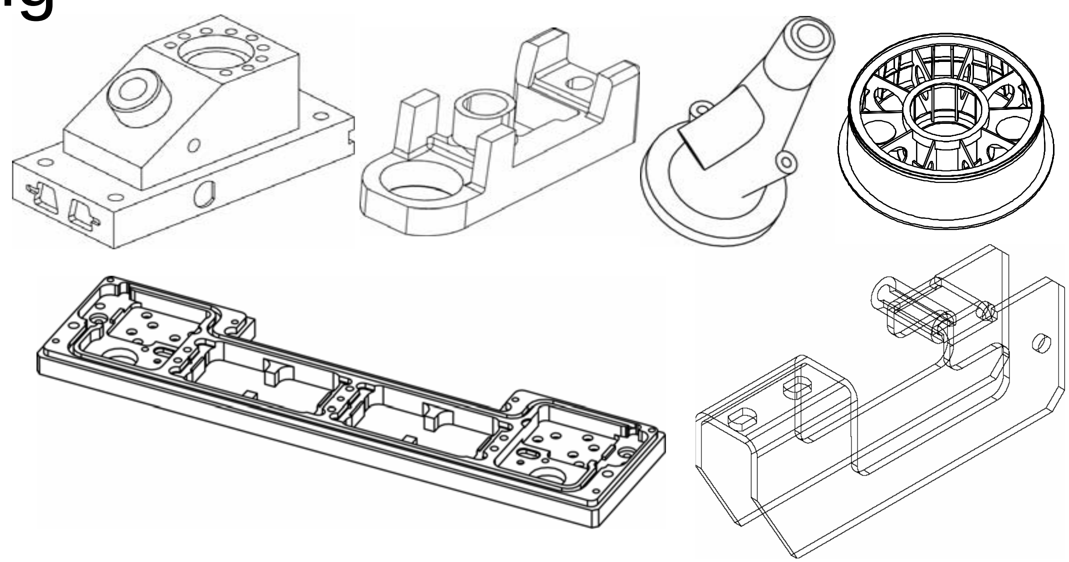
American

CAD Classifications are ***not*** Subjective



- Shape matching shares kinship with image interpretation

- CAD shares kinship with medical imaging or vision
- Attributes are rigid and unambiguous

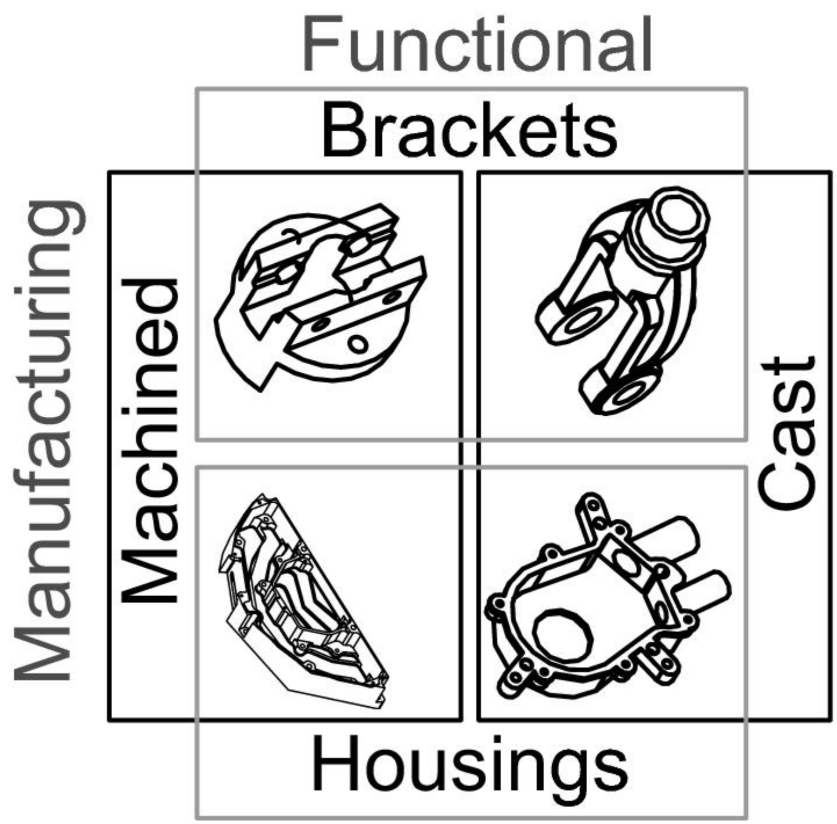


What Makes CAD Objects Different?

- Different fundamental underlying representation from graphical objects
- Engineering objects (CAD objects) have a physical realization
- Object “tags” and classifications are not subjective
- **Need to support multiple classification schemes depending on the end-user application**

Multiple Classifications

- The same parts can be classified according to different classifications.
- Example:
  - Functional:
    - Brackets vs Housings
  - Manufacturing:
    - Machined vs Cast-then-Machined



Question

- Can I make a  with just...?

3-axis machine center

Or /And

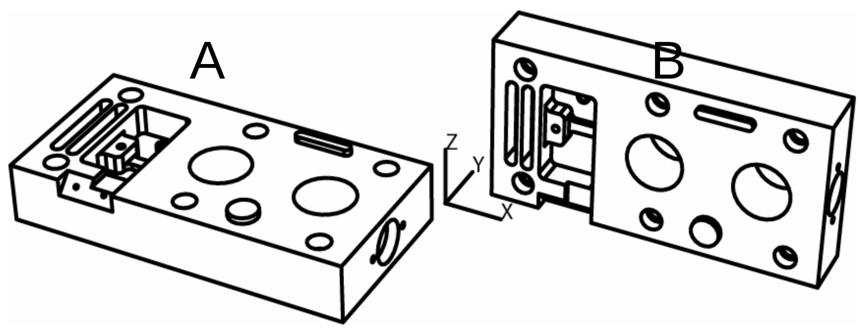
Cold chamber die casting machine

Challenges for 3D Search

1. CAD objects are different than “shape” objects
2. Focus is on the “query by example” paradigm
3. Techniques limited to individual, discrete, objects
4. Lack of well defined object semantics beyond shape
5. Lack of use cases for applications
6. No accepted procedures for how to measure performance of techniques
   - Most use their own datasets for testing
   - Evaluation procedures are sometimes opaque
   - Metrics for “success” are not standardized

Querying Design Repositories

- Find duplicates or near duplicates
  - Query-by-Example
- Applications
  - part count reduction
  - platform standardization
  - redesign for consolidation
  - cost estimation

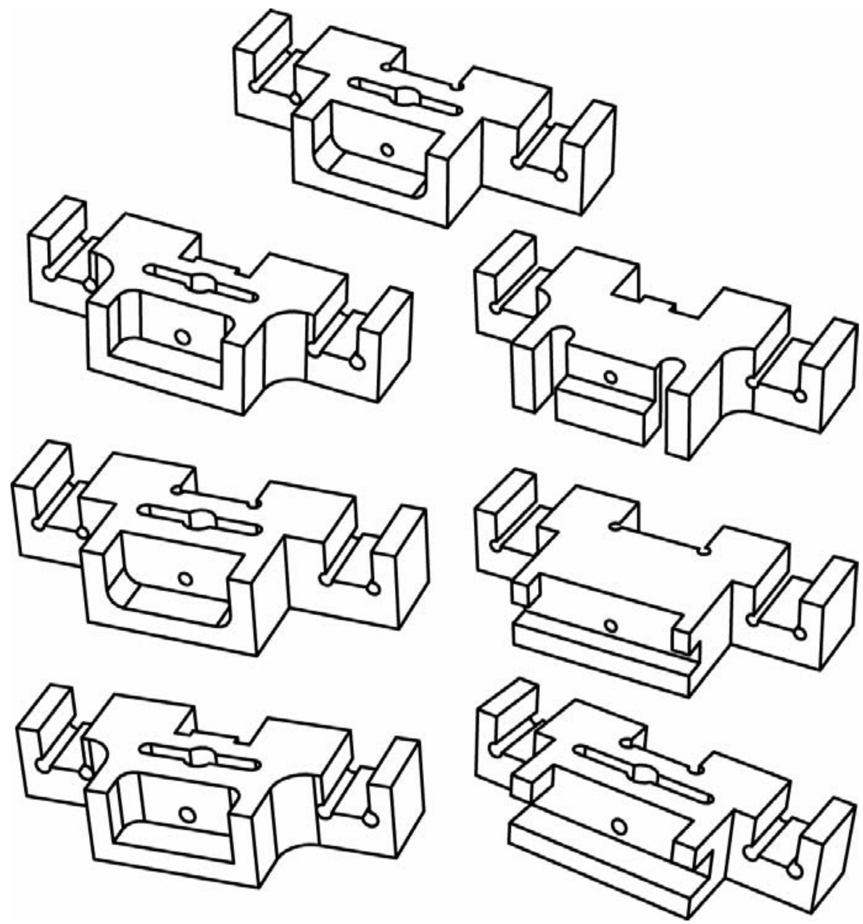


2 parts, exactly the same
different location
different orientation

No easy way to tell A==B

Querying Design Repositories

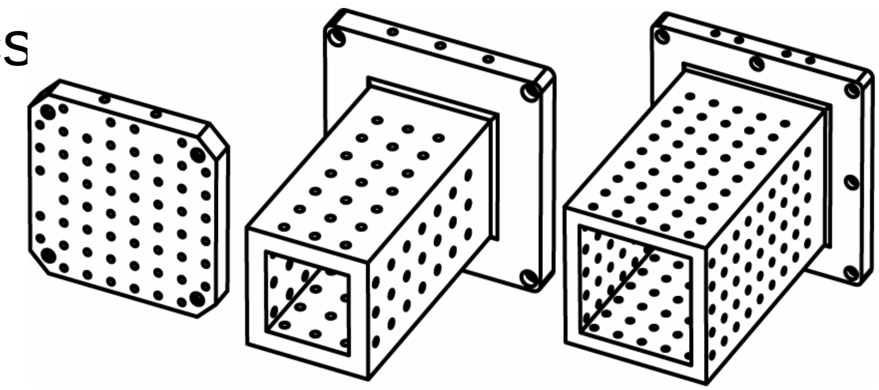
- Extract part families or design patterns
  - similar to *knowledge discovery* in large databases
- Applications
  - part count reduction
  - platform standardization
  - redesign for consolidation
  - cost estimation



a family of brackets

Querying Design Repositories

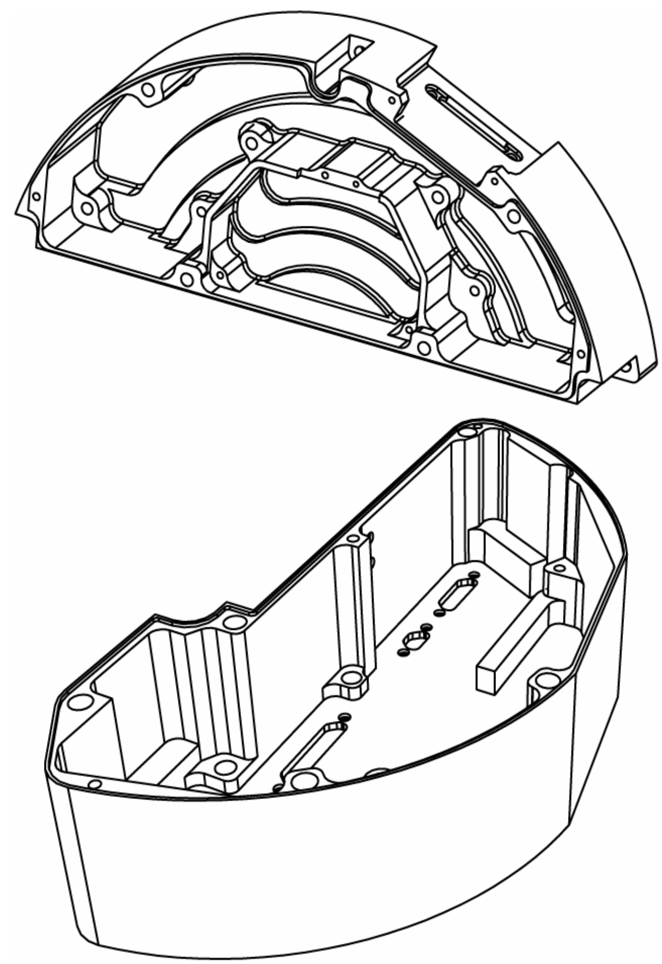
- Identify Manufacturing Clusters
  - group artifacts with similar manufacturing semantics
- Applications
  - variant process planning
  - cost estimation
  - manufacturing process & factory optimization



all holes have same *d*

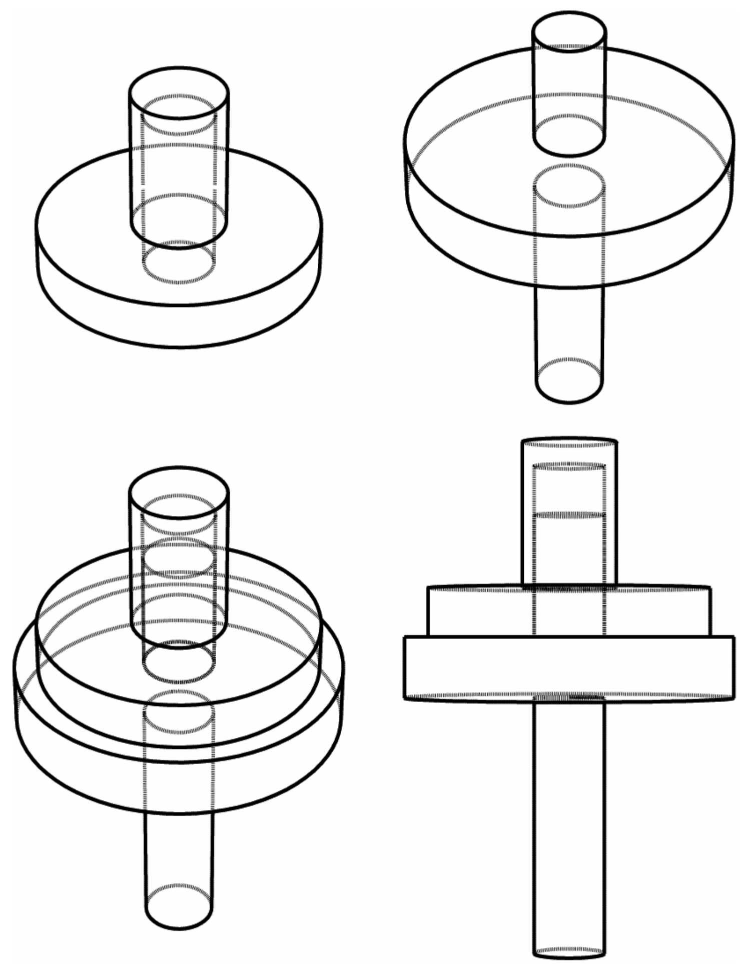
Querying Design Repositories

- Variational Design
- Variational Process Planning
- Access to corporate and institutional memory



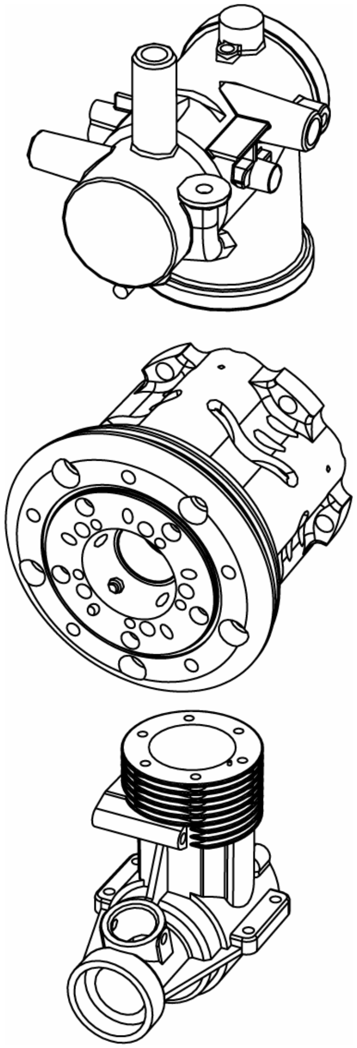
Querying Design Repositories

- Mating, geometric fit or assembly relationships
- Inverse relationships
- Constraining criteria



Querying Design Repositories

- Process Selection
  - Cluster parts based on manufacturing process criteria
  - Identify the right process for prototyping



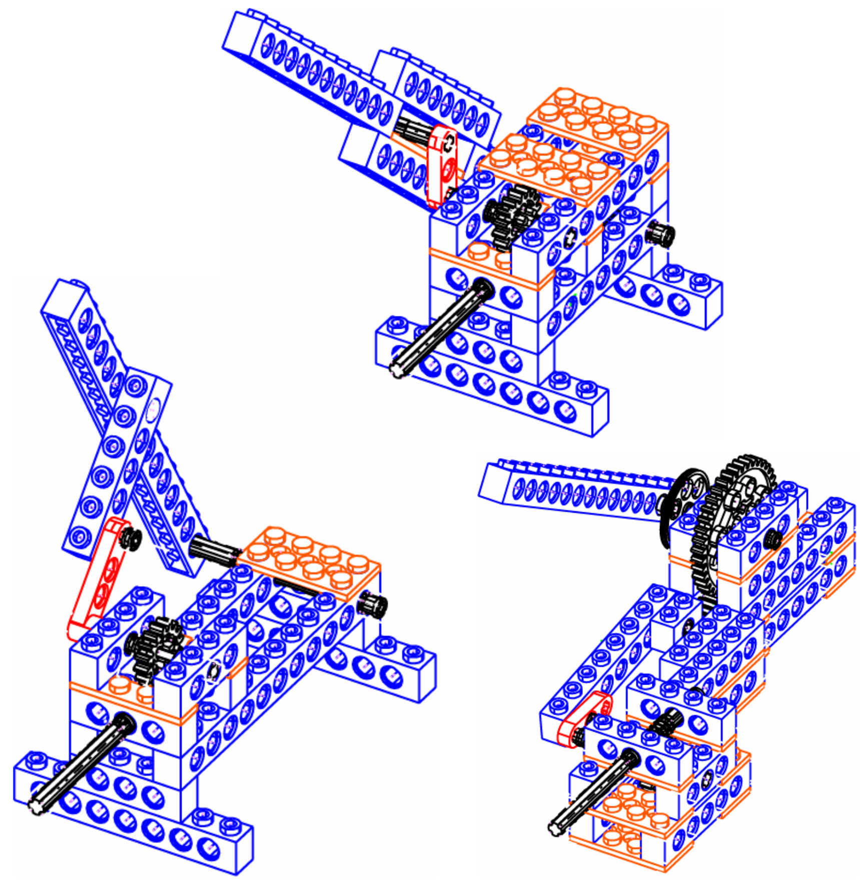
Challenges for 3D Search

1. CAD objects are different than “shape” objects
2. Focus is on the “query by example” paradigm
3. Techniques limited to individual, discrete, objects
4. Lack of well defined object semantics beyond shape
5. Lack of use cases for applications
6. No accepted procedures for how to measure performance of techniques
   – Most use their own datasets for testing
   – Evaluation procedures are sometimes opaque
   – Metrics for “success” are not standardized

Querying Design Repositories

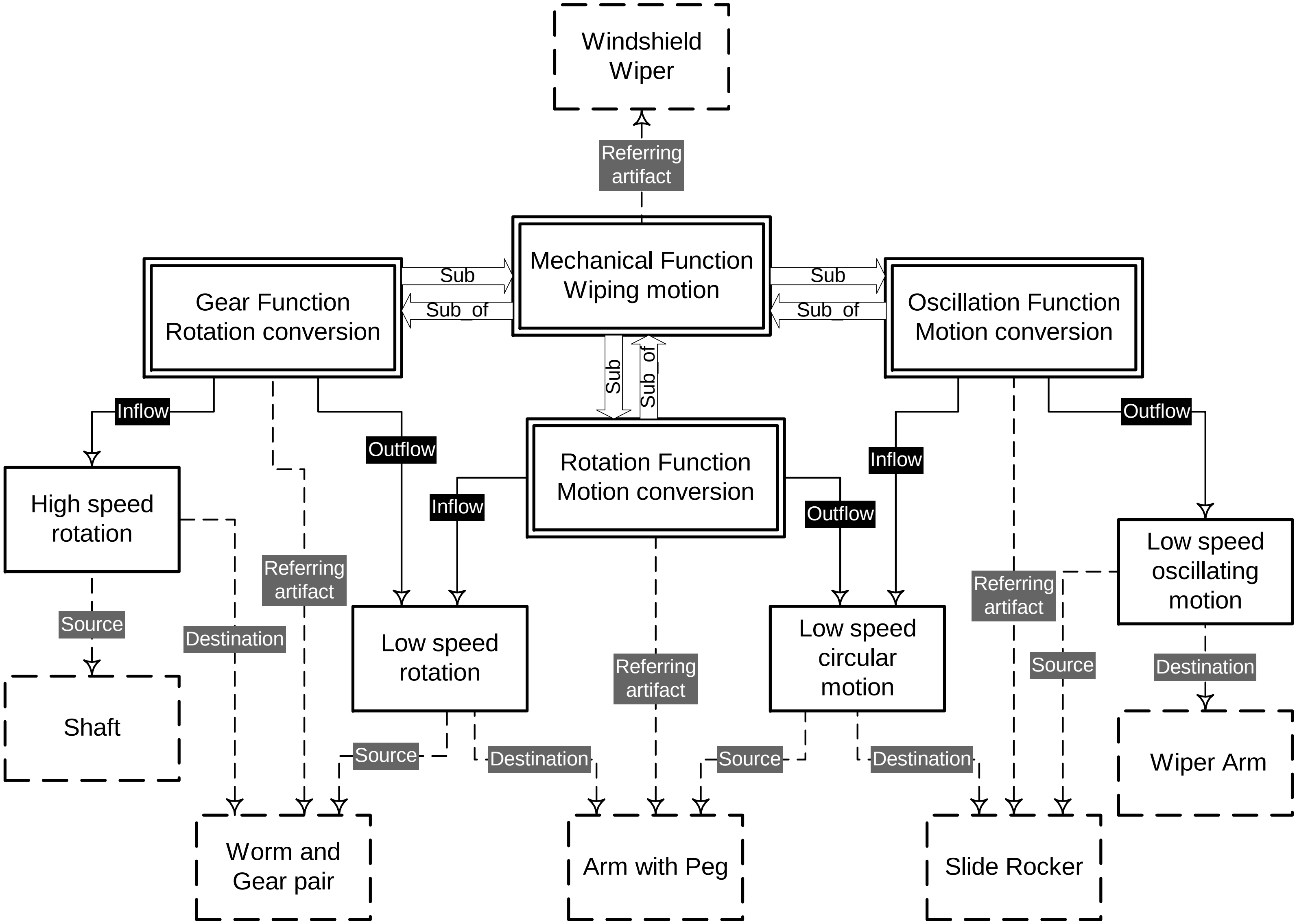
- Assembly structure, behavior and function (SBF)

- This will be much more common and economically important
- How to do the knowledge markup?



3 Lego models of a wiper assembly

Slide Rocker Windshield Wiper Design Formalised as NIST Function Flow Diagram



Challenges for 3D Search

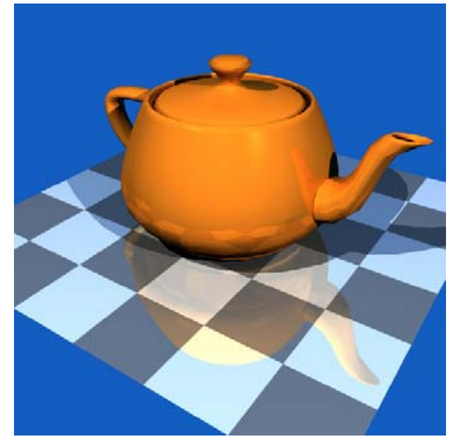
1. CAD objects are different than “shape” objects
2. Focus is on the “query by example” paradigm
3. Techniques limited to individual, discrete, objects
4. Lack of well defined object semantics beyond shape
5. Lack of use cases for applications
6. No accepted procedures for how to measure performance of techniques
   - Most use their own datasets for testing
   - Evaluation procedures are sometimes opaque
   - Metrics for “success” are not standardized

Lack of Use Cases

- Definition: “Use cases allow description of sequences of events that, taken together, lead to a system doing something useful.”
- For 3D CAD search, there is not a rich set of use cases
- Systems and business processes are already in place; new capabilities need to be inserted into existing workflow

Challenges for 3D Search

1. CAD objects are different than “shape” objects
2. Focus is on the “query by example” paradigm
3. Techniques limited to individual, discrete, objects
4. Lack of well defined object semantics beyond shape
5. Lack of use cases for applications
6. No accepted procedures for how to measure performance of techniques
   - Most use their own datasets for testing
   - Evaluation procedures are sometimes opaque
   - Metrics for “success” are not standardized



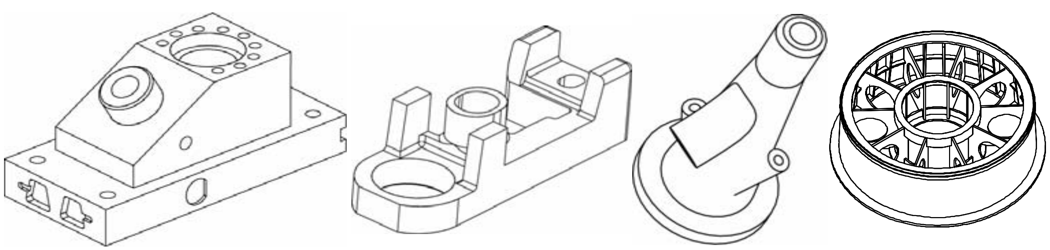
UCI Machine Learning

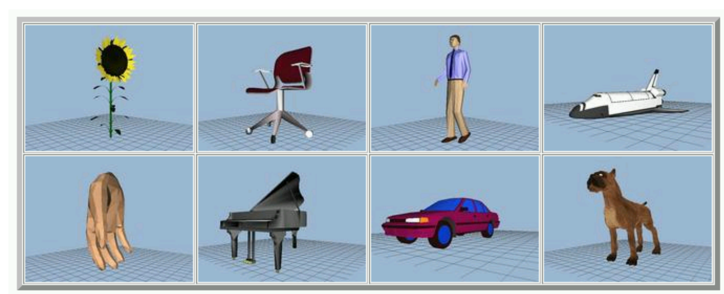
SOURCEFORGE.net

Welcome to the UCI Machine Learning Repository!

Stanford University SNLP Natural Language Processing

Statistical natural language processing and corpus-based computational linguistics: An annotated list of resources





Example: Comparing Existing Techniques

Consider the matching techniques:

1. Scale Space [this SM03 paper]
2. Scale Space w/ Decomposition Control [ASME JCISE03]
3. Shape Distributions [Osada et al ACM ToG 2002, SMI01]
4. Enhanced Shape Distributions [Ip et al SM02]
5. Learning Shape Classifiers [Ip et al SM03] (poster)
6. Matching with Reeb Graphs [Hilaga et al, SIGGRAPH 01]
7. Invariant Topology Vectors [McWherter et al, SM01]
8. Design Feature-based Matching [Cicirello et al, SM99]
9. Machining Feature-based Matching
   [Cicirello et al, SMI01, AIEDAM 2003]

- *Which* is best? W*hen* is it best?
  Why is it best?

Idea #1: The Retrieval Task

Given

- A dataset *D*
- An object *d* (possibly from *D*)
- An algorithm *A*

The Question:

Are the objects "returned" by *A* as being "similar" to *d* "really similar"?

To Measure Success: Borrow Ideas from Information Science

- Treat each object as a “document”
- Perform “Query by Example”
- Measure *recall* and *precision*
  - Recall: % of relevant documents retrieved relative to the entire set of relevant documents
  - Precision: % of relevant documents retrieved relative to all those retrieved
- Ideal: 100% recall, 100% precision

Problems with this Approach

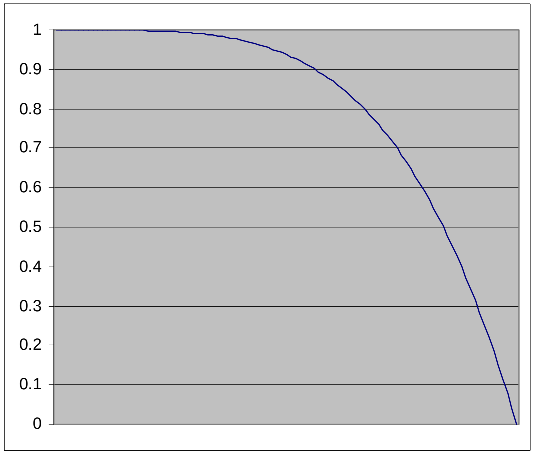
- Really intended for document retrieval
  - Text searching
- Requires an a priori labeling scheme
  - Such a labeling is nearly always done by humans and can be highly subjective
- Assumes an absolute labeling scheme
  - If the document contains the keyword “Spline” it has to do with Splines
- Labeling are cumulative
  - Documents might be about “graphics + Spline” or “shipbuilding + spline” etc, the fact they share spline will make them similar

Further IR doesn't work for CAD-based Engineering Data

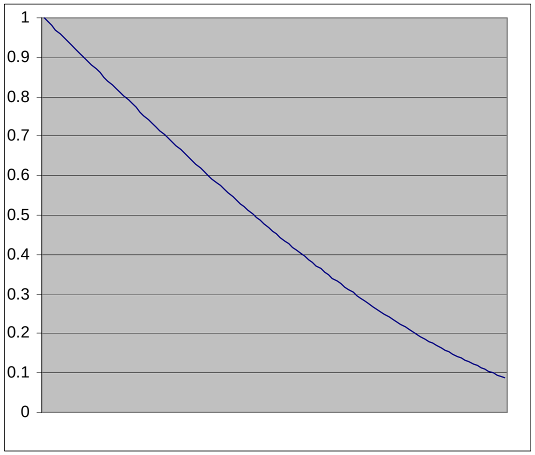
- CAD models are not consistent documents
  - "Text" is in the eye of the beholder
  - E.g. manufacturing view different from design view
- No agreed upon labeling scheme
  - Features vary by domain, can't manually classify 1,000,000 parts! (this was GT's problem…)
- Labeling schemes are not absolute
  - At the geometry level, similar shape features could still have different functionally or mfg. properties
- Labeling are not cumulative
  - Labels are orthogonal; objects under "housing + hole" are not going to be at all similar to "bolt + hole"

Note: IR doesn't seem to work for shape models either…

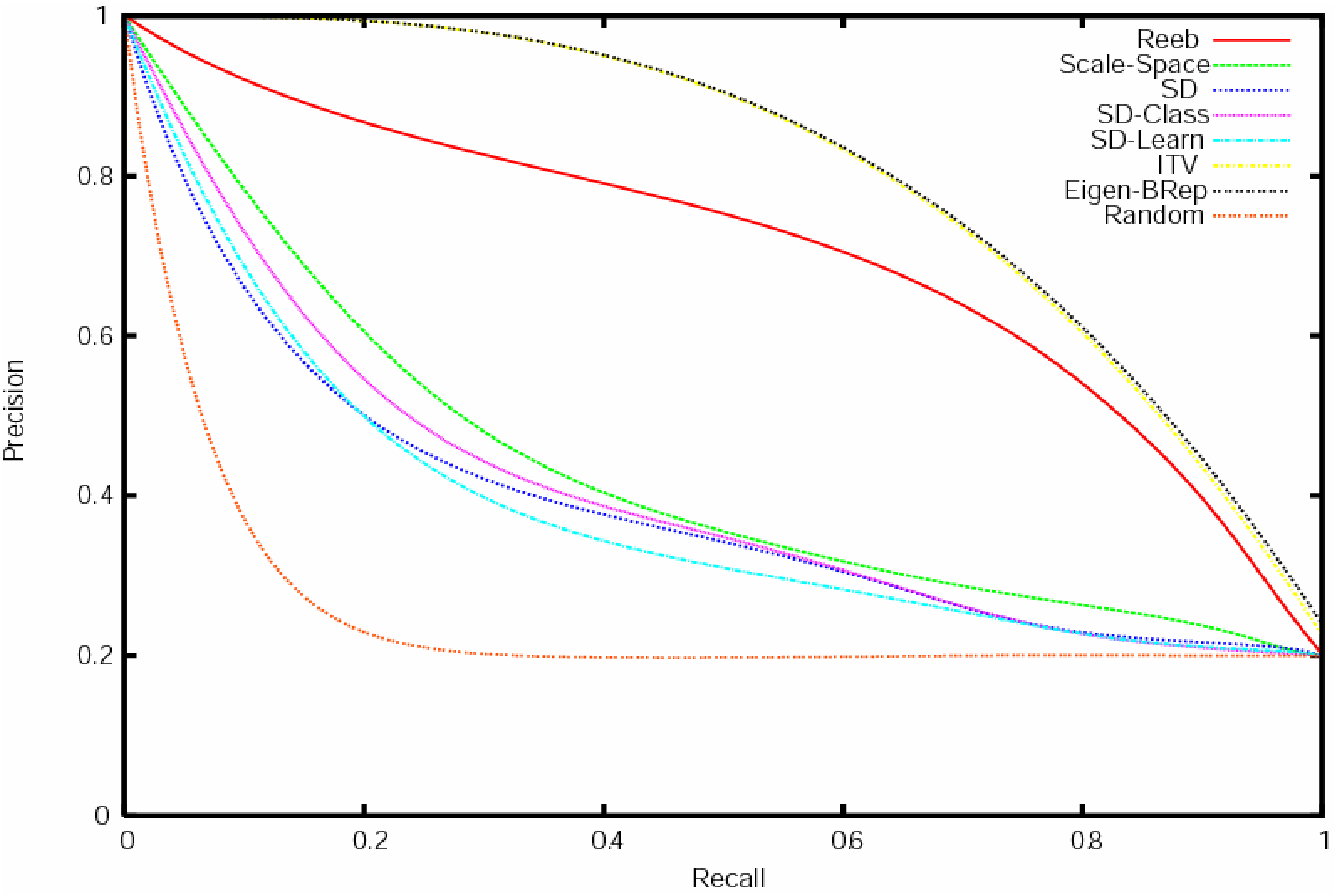
- Existing results are the opposite of what one wants
- What we want



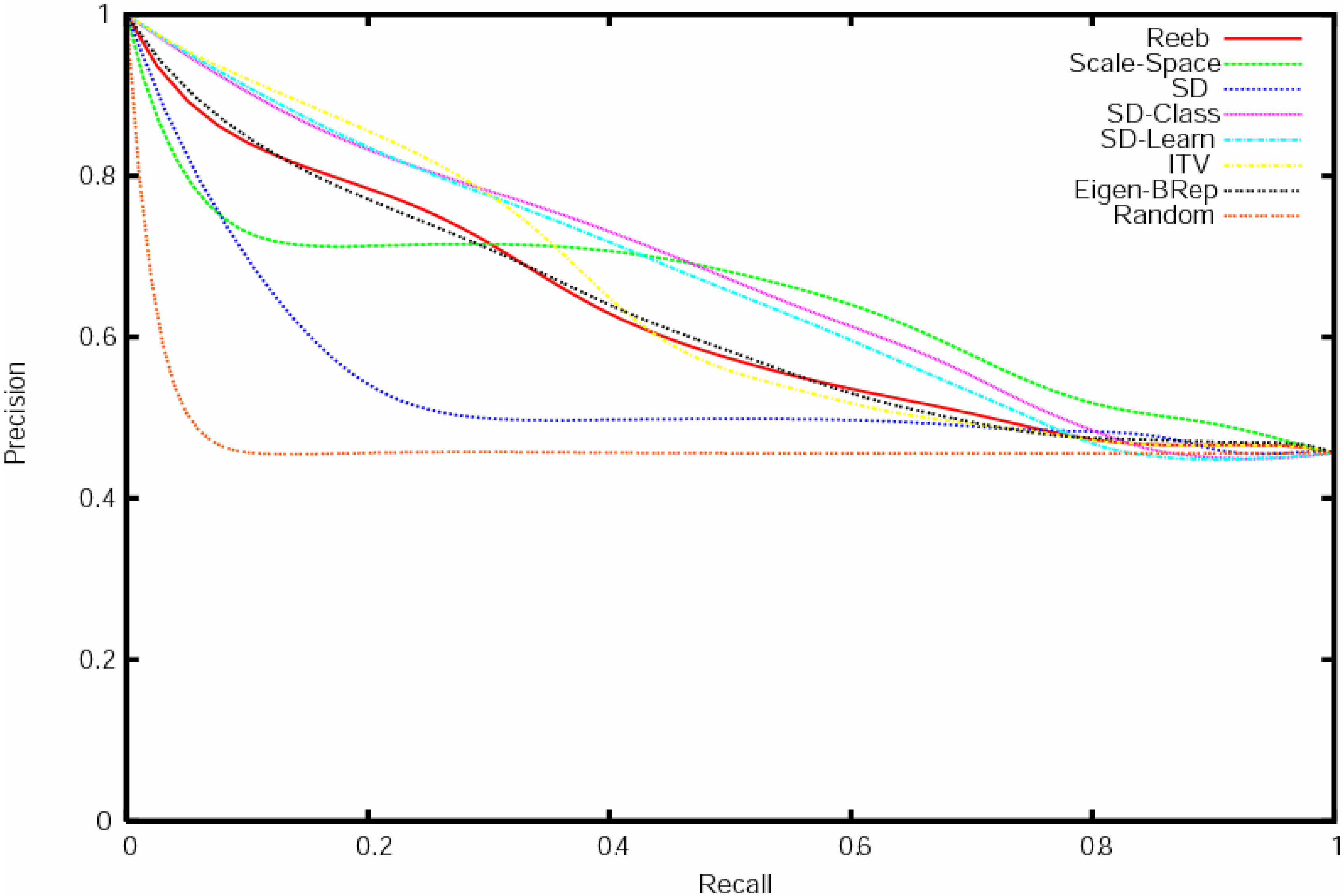
- What we get



Minor Topological Variation Dataset: Cubes



LEGO Dataset



Retrieval may be the wrong problem to solve…

- These matching algorithms actually implicitly encode *shape classifiers* (Kyprianou's goal)
- <u>Classification</u>
  - – Assess the *quality* of the classification achieved
  - – against the desired (a priori) classification
- There is a scientific justification & motivation for this idea in *statistical learning theory* and *machine learning*

Idea #2: The Classification Task

Given

- a dataset D
- a classification of D
  $C = \{c_1\ c_2\ c_3\ \ldots\ c_k\}$
  - c_i's are a partition of D
- an algorithm A

The Question:

How well does A do in reproducing classification C ?

The Classification Task: Algorithm Selection

Given

- a dataset D
- a classification of D
  $C = \{c_1\ c_2\ c_3\ \ldots\ c_k\}$
  - c_i's are a partition of D
- a set of algorithms $A_1, A_2, \ldots A_n$

The Question:
Which A_i is best at reproducing this classification?

To Measure Success: Borrow Ideas from Machine Learning

Treating

- the data set as unseen data
- the algorithms, A_1, A_2, … A_n , as pre-trained classifiers
- the classification as the error function
- And using information gain and error measures to calculate which is best for the dataset

An Example

Compare

1. Scale Space [this SM03 paper]
2. Scale Space w/ Decomposition Control [ASME JCISE03]
3. Shape Distributions [Osada et al ACM ToG 2002, SMI01]
4. Enhanced Shape Distributions [Ip et al SM02]
5. Learning Shape Classifiers [Ip et al SM03] (poster)
6. Matching with Reeb Graphs [Hilaga et al, SIGGRAPH 01]
7. Invariant Topology Vectors [McWherter et al, SM01]

On CAD_40 dataset

Shape Classification Quality

G = Group
FP = False-Positives
FN = False-Negatives

Scale-Space

Total Errors: 15

G1: FP: 0, FN: 3	G6: FP: 1, FN: 1
G2: FP: 0, FN: 2	G7: FP: 1, FN: 2
G3: FP: 1, FN: 1	G8: FP: 2, FN: 1
G4: FP: 1, FN: 1	G9: FP: 5, FN: 1
G5: FP: 1, FN: 3	G10:FP: 3, FN: 0

Scale-Space with automated feature extraction

Total Errors: 12

G1: FP: 0, FN: 3	G6: FP: 0, FN: 1
G2: FP: 0, FN: 1	G7: FP: 1, FN: 1
G3: FP: 2, FN: 1	G8: FP: 0, FN: 1
G4: FP: 1, FN: 1	G9: FP: 4, FN: 1
G5: FP: 1, FN: 2	G10:FP: 3, FN: 0

Multi-resolution Reeb Graphs

Total Error: 13

G1: FP: 0, FN: 3	G6: FP: 3, FN: 1
G2: FP: 0, FN: 1	G7: FP: 1, FN: 1
G3: FP: 0, FN: 1	G8: FP: 0, FN: 1
G4: FP: 0, FN: 2	G9: FP: 4, FN: 1
G5: FP: 1, FN: 2	G10:FP: 4, FN: 0

Original shape distributions

Total Errors: 11

G1: FP: 0, FN: 2	G6: FP: 0, FN: 1
G2: FP: 0, FN: 2	G7: FP: 1, FN: 1
G3: FP: 1, FN: 1	G8: FP: 3, FN: 1
G4: FP: 3, FN: 1	G9: FP: 1, FN: 1
G5: FP: 0, FN: 1	G10:FP: 2, FN: 0

Enhanced distributions

Total Errors: 14

G1: FP: 0, FN: 5	G6: FP: 0, FN: 2
G2: FP: 0, FN: 1	G7: FP: 0, FN: 1
G3: FP: 1, FN: 1	G8: FP: 1, FN: 1
G4: FP: 6, FN: 1	G9: FP: 2, FN: 1
G5: FP: 0, FN: 1	G10:FP: 4, FN: 0

Enhanced distributions + Learning

Total Errors: 13

G1: FP: 0, FN: 4	G6: FP: 0, FN: 1
G2: FP: 0, FN: 2	G7: FP: 0, FN: 1
G3: FP: 1, FN: 1	G8: FP: 2, FN: 1
G4: FP: 5, FN: 1	G9: FP: 1, FN: 1
G5: FP: 1, FN: 1	G10:FP: 3, FN: 0

ITV topology

Total Errors: 15

G1: FP: 0, FN: 2	G6: FP: 1, FN: 2
G2: FP: 0, FN: 2	G7: FP: 1, FN: 1
G3: FP: 1, FN: 1	G8: FP: 2, FN: 1
G4: FP: 1, FN: 1	G9: FP: 7, FN: 2
G5: FP: 0, FN: 3	G10:FP: 2, FN: 0

Comparison Results

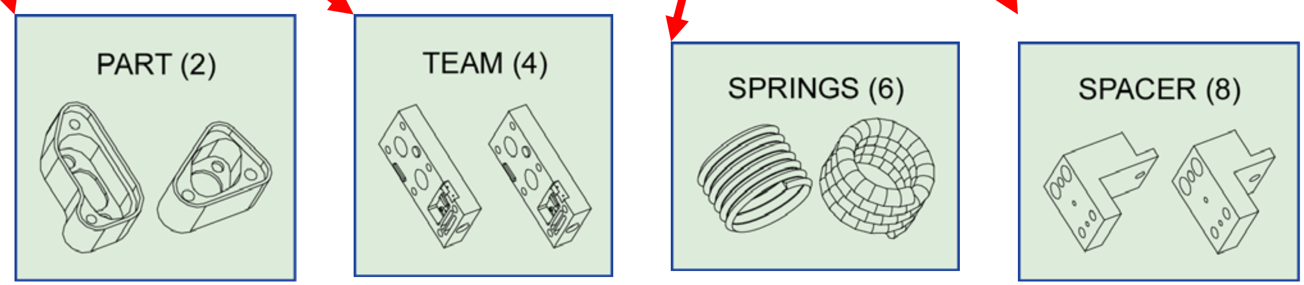
Finding best performance (by model group):

G1: Reeb Graph and ITV	**G6: Scale-Space, Orig. and Learn. Dist.**
G2: Scale-Space, Orig. and Enh. Dist.	**G7: Enh. and Learn. Dist.**
G3: Reeb Graph	**G8: Scale-Space and Reeb Graph**
G4: Scale-Space, Reeb Graph, ITV	**G9: Orig. and Learn. Dist.**
G5: Orig. and Enh. Dist.	**G10: Orig. Dist and ITV**

Comparison Results

Finding best performance (by model group):

G1: Reeb Graph and ITV
G2: Scale-Space, Orig. and Enh. Dist.
G3: Reeb Graph
G4: Scale-Space, Reeb Graph, ITV
G5: Orig. and Enh. Dist.
G6: Scale-Space, Orig. and Learn. Dist.
G7: Enh. and Learn. Dist.
G8: Scale-Space and Reeb Graph
G9: Orig. and Learn. Dist.
G10: Orig. Dist and ITV

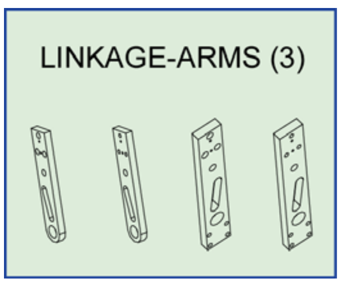


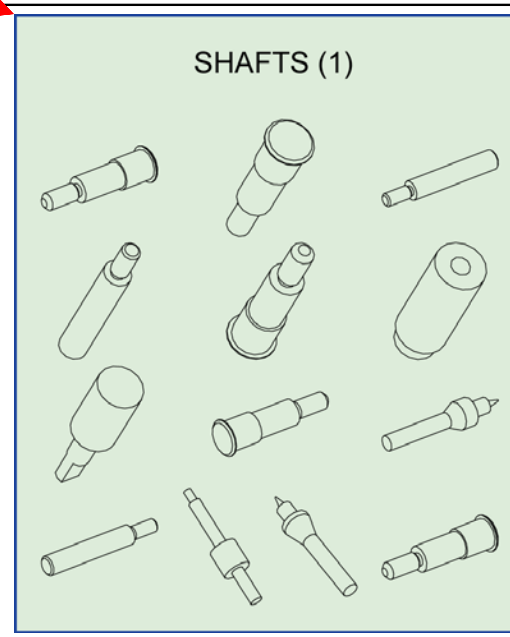
Why?: feature decompositions create nearly identical trees.

Comparison Results

Finding best performance (by model group):

G1: Reeb Graph and ITV	**G6: Scale-Space, Orig. and Learn. Dist.**
G2: Scale-Space, Orig. and Enh. Dist.	**G7: Enh. and Learn. Dist.**
G3: Reeb Graph	**G8: Scale-Space and Reeb Graph**
G4: Scale-Space, Reeb Graph, ITV	**G9: Orig. and Learn. Dist.**
G5: Orig. and Enh. Dist.	**G10: Orig. Dist and ITV**



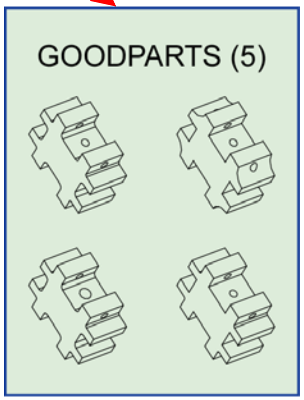


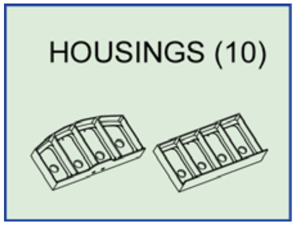
Why?: Classes are topologically homogeneous.

Comparison Results

Finding best performance (by model group):

G1: Reeb Graph and ITV
G2: Scale-Space, Orig. and Enh. Dist.
G3: Reeb Graph
G4: Scale-Space, Reeb Graph, ITV
G5: Orig. and Enh. Dist.
G6: Scale-Space, Orig. and Learn. Dist.
G7: Enh. and Learn. Dist.
G8: Scale-Space and Reeb Graph
G9: Orig. and Learn. Dist.
G10: Orig. Dist and ITV



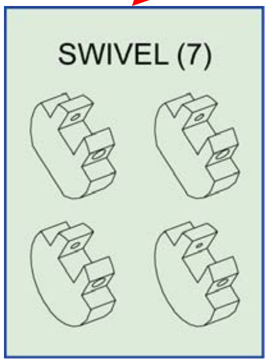


Why?: Models have strong geometric regularities.

Comparison Results

Finding best performance (by model group):

G1: Reeb Graph and ITV	**G6: Scale-Space, Orig. and Learn. Dist.**
G2: Scale-Space, Orig. and Enh. Dist.	**G7: Enh. and Learn. Dist.**
G3: Reeb Graph	**G8: Scale-Space and Reeb Graph**
G4: Scale-Space, Reeb Graph, ITV	**G9: Orig. and Learn. Dist.**
G5: Orig. and Enh. Dist.	**G10: Orig. Dist and ITV**



Why?: Class identification requires factoring geometry, topology and dimension.

Discussion of Comparison Results

- Interesting, but …..
  - Any differences in overall error rates are not statistically significant (e.g. 11-15 False Positives)
  - We need several datasets with more (>30) classes and (perhaps) larger (>30) numbers of classes
  - We need standard data sets and test cases!
- Observations:
  - Data quality can influence the results
  - Performance varied greatly by model class… But some are *clearly better* for certain classes
  - No one technique is better overall

The General Classification Algorithm Selection Task

Given

- a representative training dataset D
- a classification of D
  $C = \{c_1\ c_2\ c_3\ \dots\ c_k\}$
  - c_i's are a partition of D
- A set of algorithms $A_1, A_2, \dots A_n$

The Question:

Which A_i is best at classifying unseen objects (probably from a class like D) in a manner consistent with C ?

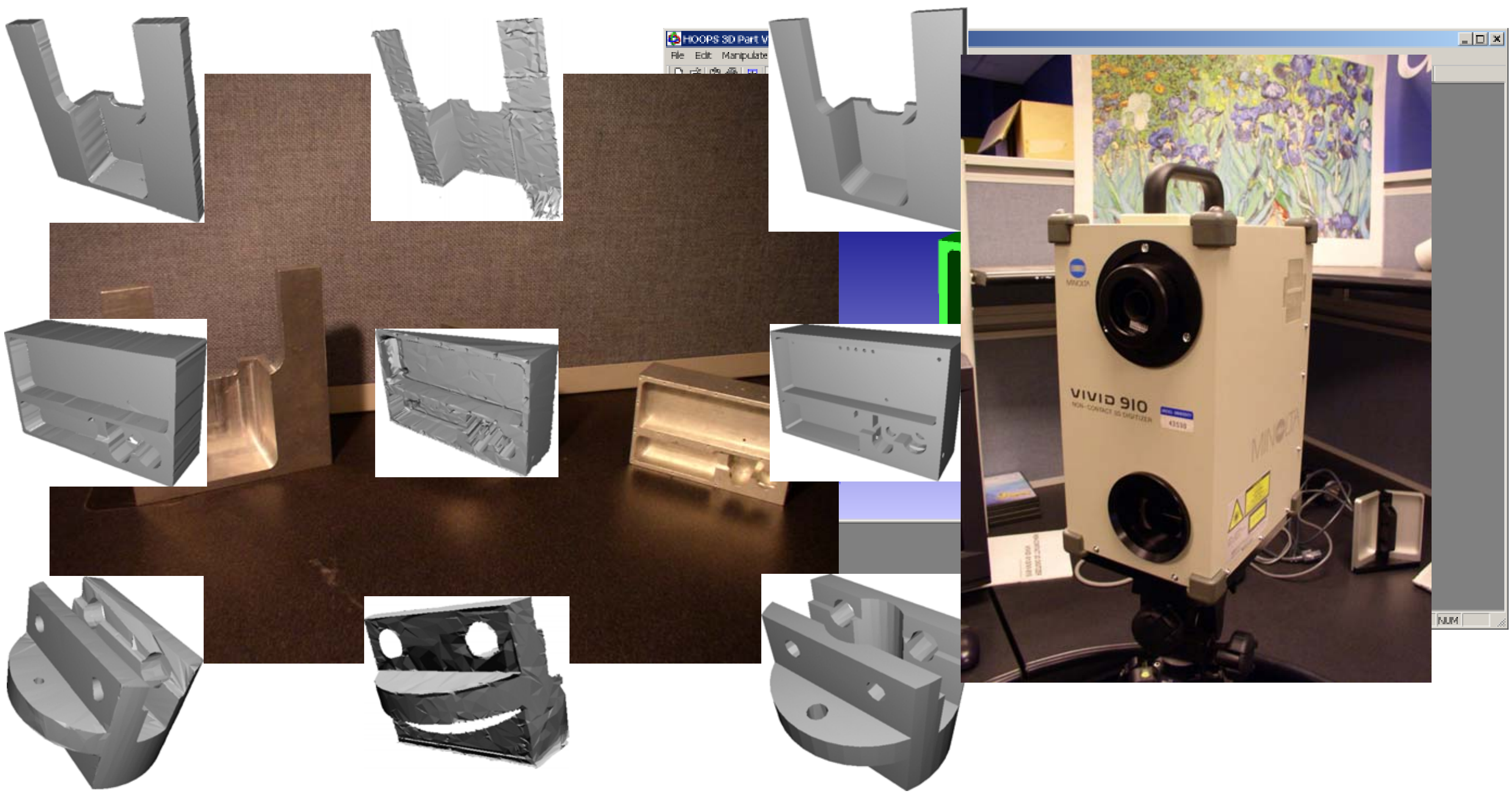
Discussion Points…

- No technique is clearly better; most perform poorly
- Is there something better than Precision-Recall?
- More work needed on how to better use:
  - Boundary representations
  - Feature-based techniques
- Which engineering questions to answer?
- Answering engineering questions is challenging
  - Manufacturing classifications & functional classifications
  - We need better specifications on engineering questions
- Datasets need to be bigger and more widely available

A CAD Search Frontier:
Partial Matching of Acquired Data

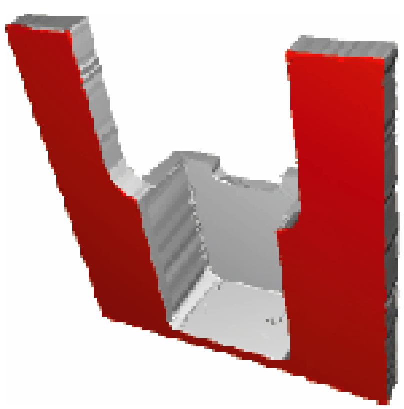
Experimental Results – Scanned Data

360° Scan | Single Scan | From Exact Representation

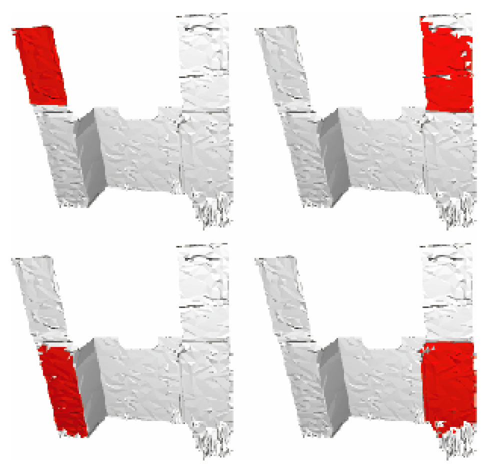


Experimental Results – Scanned Data

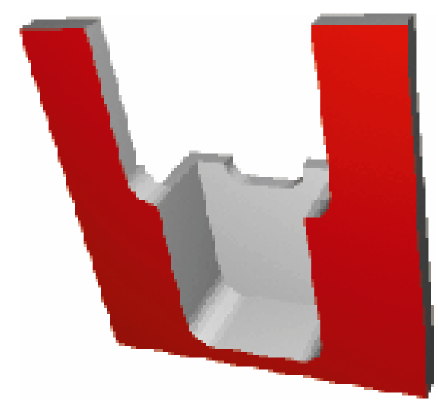
360° Scan



Single Scan



From Exact Representation

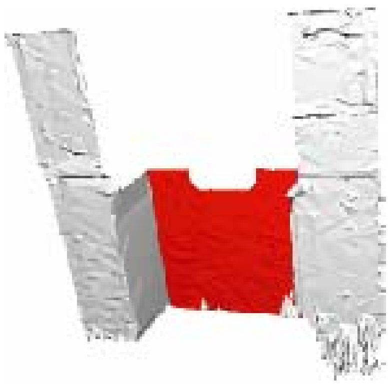


An example of one-to-many correspondence

Experimental Results – Scanned Data

360° Scan

Single Scan

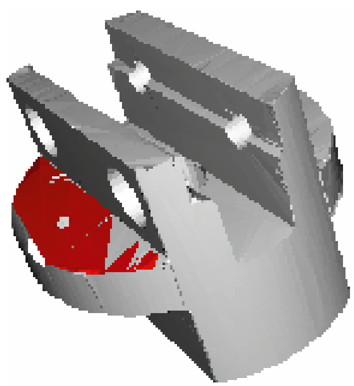


From Exact Representation

An example of one-to-one correspondence

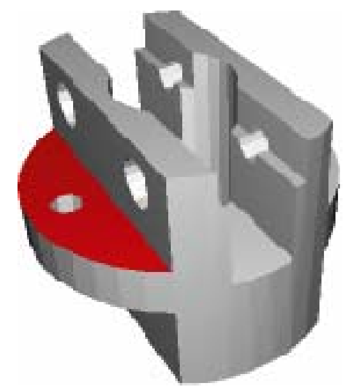
Experimental Results – Scanned Data

360° Scan



Single Scan

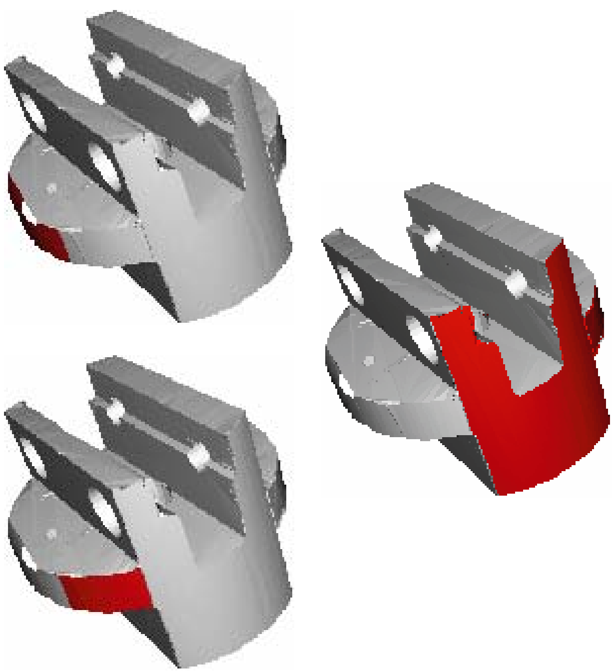
From Exact Representation



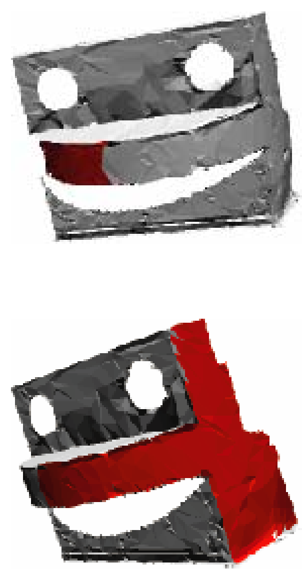
An example of one-to-one correspondence

Experimental Results – Scanned Data

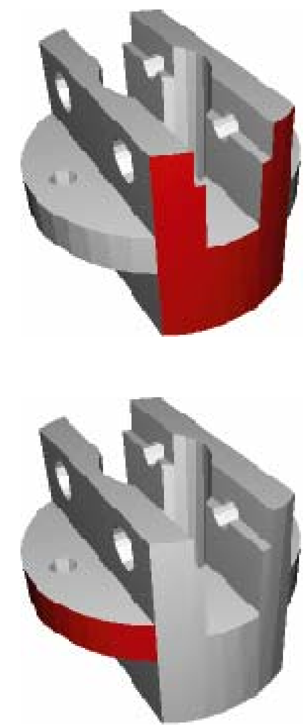
360° Scan



Single Scan



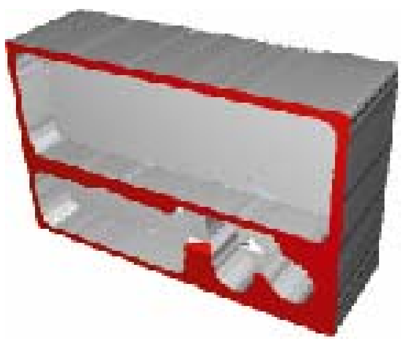
From Exact Representation



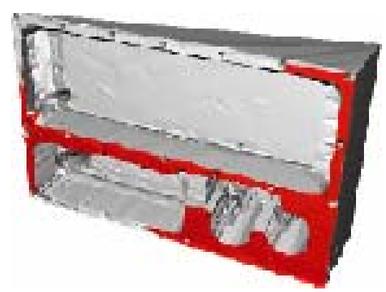
An example of many-to-many correspondence

Experimental Results – Scanned Data

360° Scan



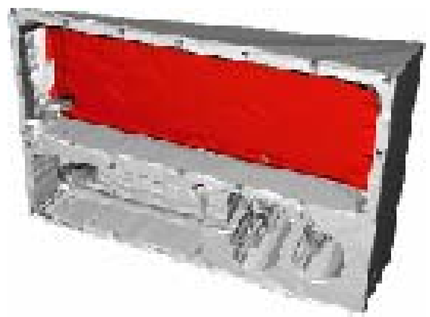
Single Scan



From Exact Representation

An example of one-to-many correspondence

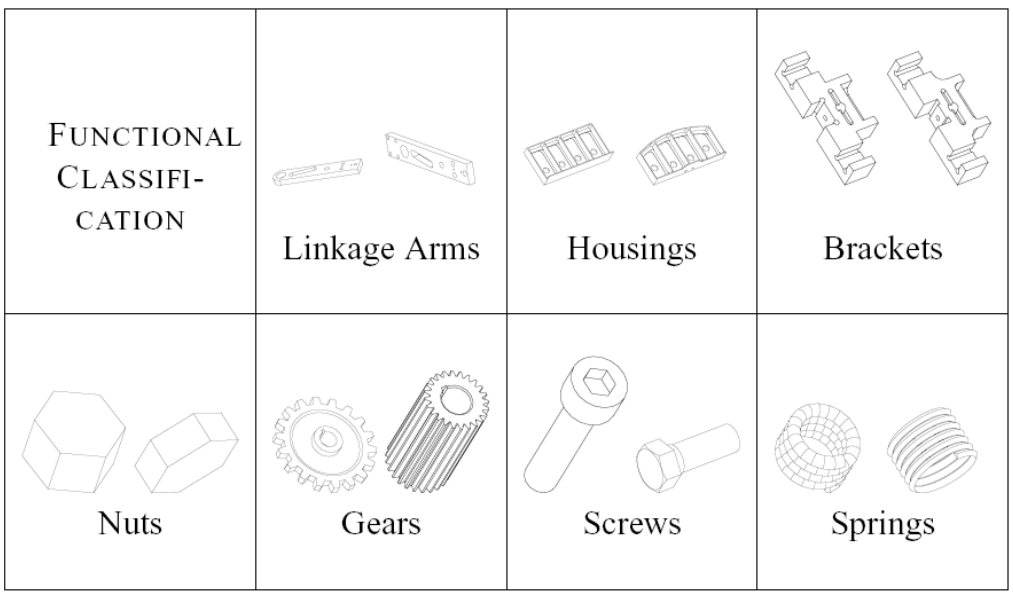
Experimental Results – Scanned Data

360° Scan	Single Scan	From Exact Representation
		

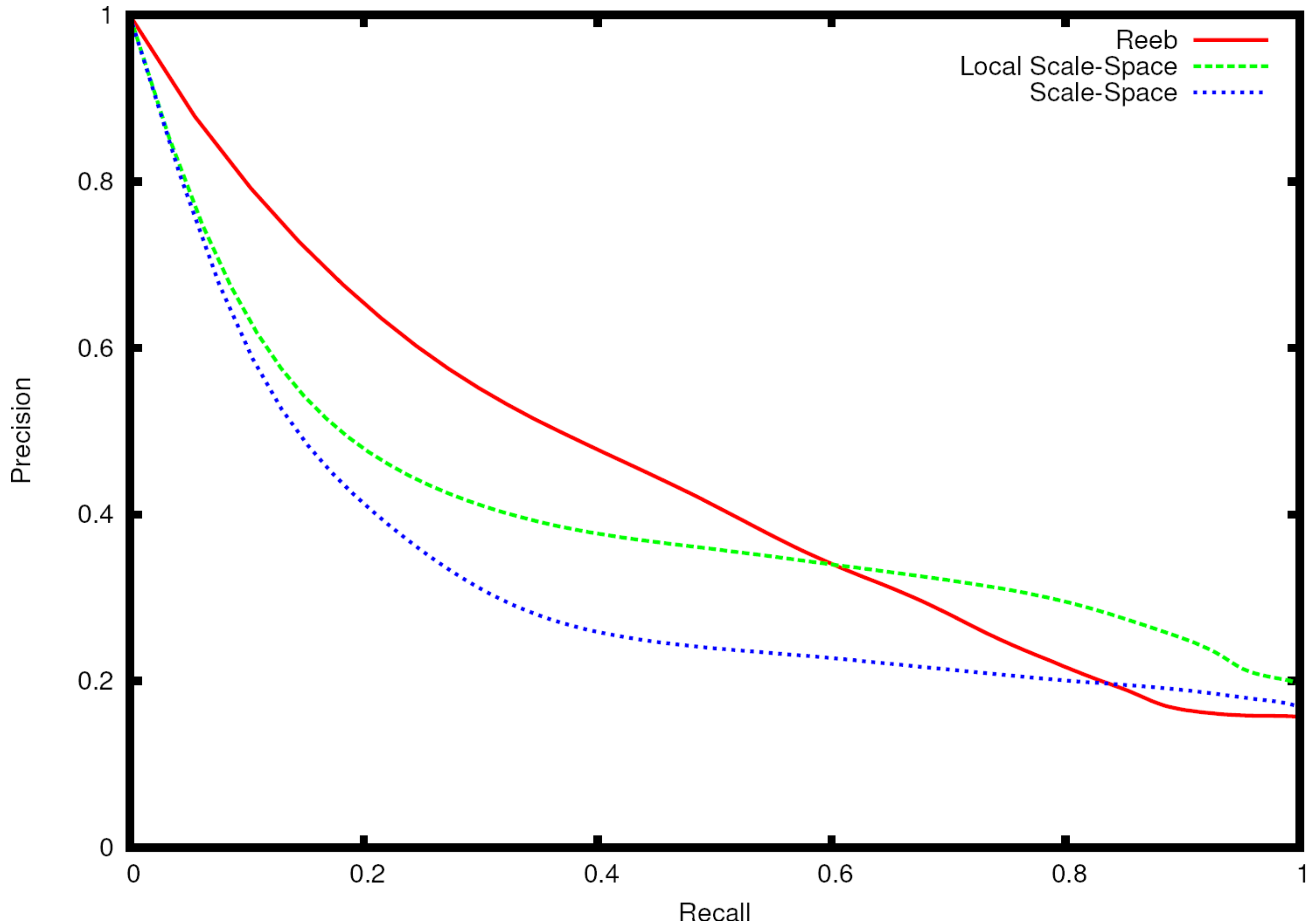
An example of one-to-one correspondence

Retrieval Using Functional Classification

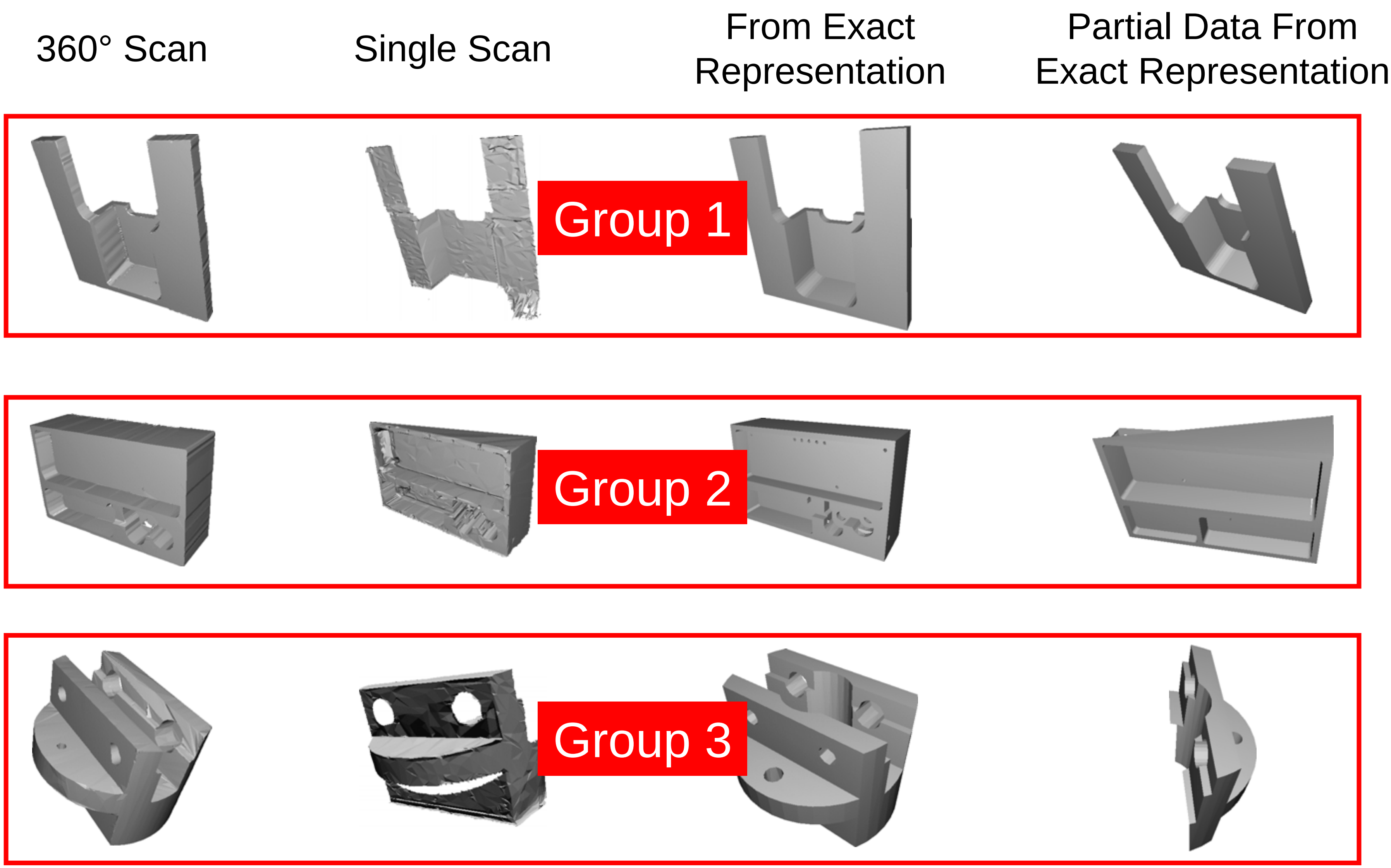
- Techniques used:
  - Reeb graph comparison (Reeb)
  - Global Scale-Space (Scale-Space)
  - Local Scale-Space (Local Scale-Space)



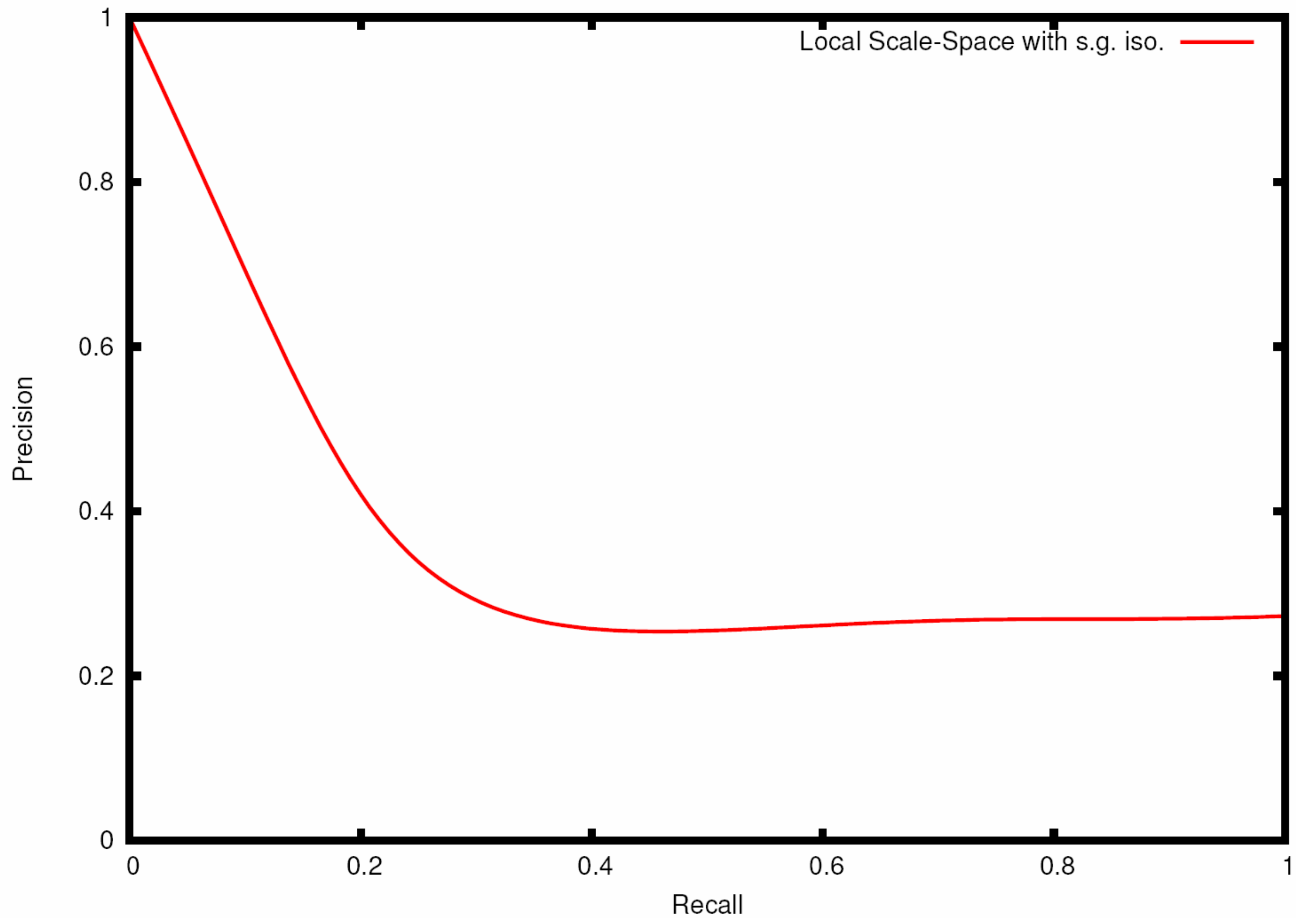
Retrieval Using Functional Classification



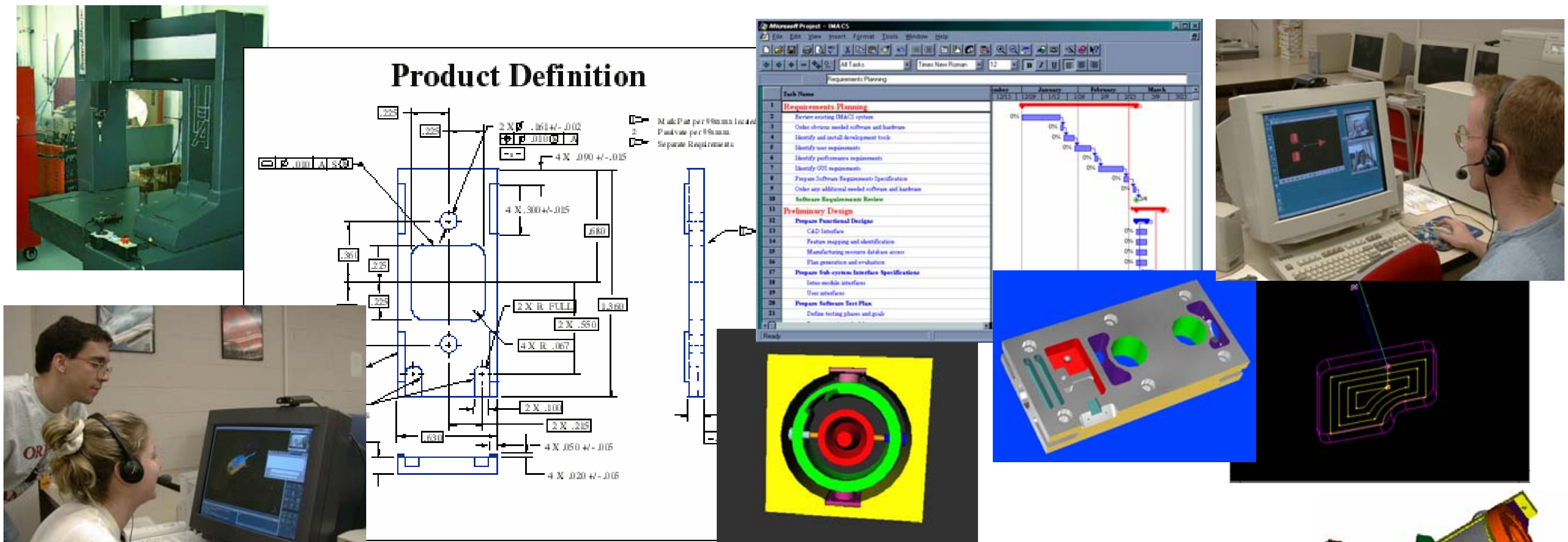
Retrieval on Partial and Scanned Data



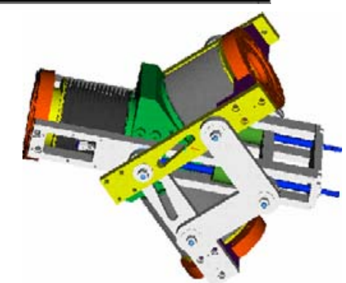
Retrieval on Partial and Scanned Data



Design Repositories:
Digital Libraries for Engineering



$$\tau_s = \mathbf{M}(\boldsymbol{\theta})\cdot\ddot{\boldsymbol{\theta}} + \mathbf{V}(\boldsymbol{\theta},\dot{\boldsymbol{\theta}}) + \mathbf{G}(\boldsymbol{\theta}) + \mathbf{F}(\boldsymbol{\theta},\dot{\boldsymbol{\theta}})$$



Engineering Digital Libraries with CAD models, assemblies, process plans, revisions, S-B-F models, project information and workflows, design rationale, design history, records of collaborative activity...

Q&A

For more information

http://gicl.cs.drexel.edu

http://www.designrepository.org

National Science Foundation (NSF)

Digital Archiving and Long-Term Preservation (DIGARCH) Award NSF CISE/IIS-0456001

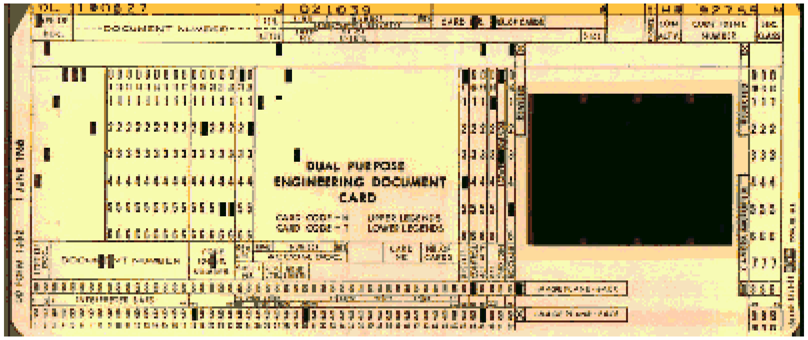
Cyber-Infrastructure Teams Awards OCI-0636235, OCI-0636273, SCI-0537125 and SCI-0537370

Observations

- No technique is clearly better; most perform poorly
- Is there something better than Precision-Recall?
- More work needed on how to better use:
  - Boundary representations
  - Feature-based techniques
- Answering engineering questions is challenging
  - Manufacturing classifications & functional classifications
  - We need better specifications on engineering questions
- Work in this paper is *completely reproducible*
  - http://edge.mcs.drexel.edu/repository/datasets/
  - If you think you can do better, go for it!
- Datasets need to be bigger

Conclusions

- Establish datasets for evaluating retrieval techniques on realistic CAD/CAM artifacts
- Describe general benchmarking procedure
  - This procedure can be followed by others
    - Please suggest improvements!
- Benchmarked nine different 3D shape and solid model matching techniques



An Approach to the General Algorithm Selection Task

First:

- Given
  - a shape matching algorithm, *A*
  - a dataset *D*
  - a classification, *C,* of *D*
- Compute the *intrinsic classifier* for *A*
- Use the *intrinsic classifier* to compute the optimal classification possible for *D,* C_o
- Find correspondence between *C* and C_o
- Compute total type1 and type2 errors between *C* and C_o

An Approach to the General Algorithm Selection Task

Second:

- Do this over all your candidate algorithms, $A_1, A_2, \ldots A_n$
- Compute relative error rates (information gain) among these algorithms
- Select the algorithm with highest information content

Observations

- This is an objective approach
  - No biases
- It will work for any matching algorithm or classification scheme, including those based on subjective human labels
- It will work regardless of the type of data
  - CAD data, shape data, bio-med data, legos, etc.
- Pending submission to SM04

Consider the B-52