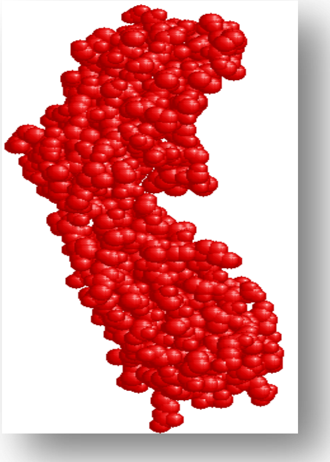
Lecture 13

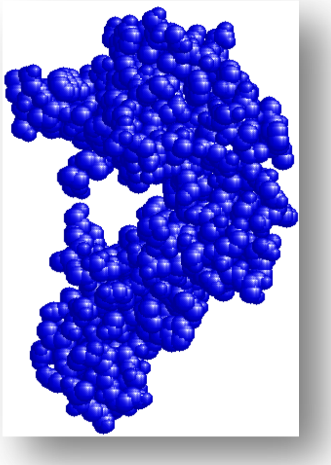
Protein-protein Docking

Protein-Protein Docking

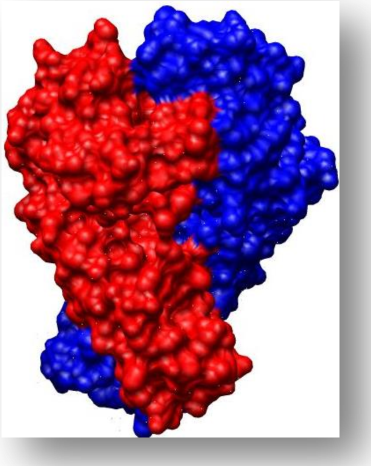
Ab initio **Protein-protein docking** is the determination of the molecular structure of *complexes* formed by two or more proteins without the need for *experimental* measurement.



+



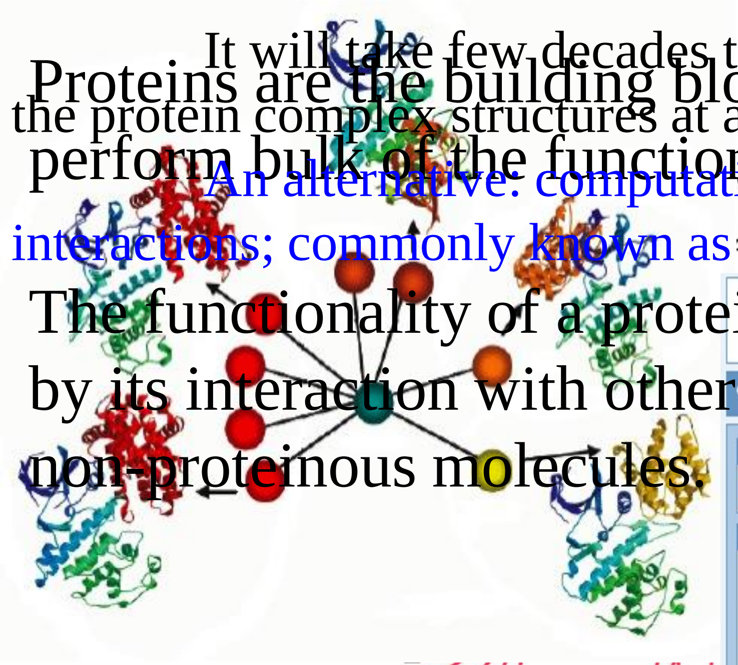
=



Background

Proteins are the building blocks of the cells; perform bulk of the functions of the cell.

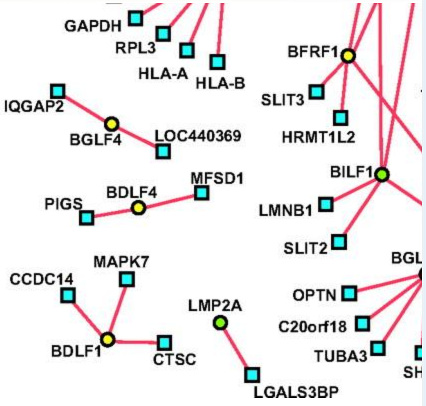
The functionality of a protein is determined by its interaction with other proteinous or non-proteinous molecules.



Aloy et al. (2004). *Nat. Biotechnology*

It will take few decades to experimentally determine all the protein complex structures at atomic level resolution.

An alternative: computational modeling of protein-protein interactions; commonly known as protein-protein docking.



Epstein–Barr virus(

A MEMBER OF THE PDB

An Information Portal to Biological Macromolecular Structures

As of Tuesday Aug 16, 2011 at 5 PM PDT there are 75246 Structures | PDB Statistics

PROTEIN DATA BANK

PDB ID or Text | PDB ID lookup or Text search of the complete structure file | Search | Advanced Search

Customize This Page

MyPDB — Hide
Login to your Account
Register a New Account

Home — Hide
News & Publications
Usage/Reference Policies
Deposition Policies
Website FAQ
Deposition FAQ
Contact Us
About Us
Careers
External Links
Sitemap
New Website Features

Deposition — Hide
All Deposit Services
Electron Microscopy
X-ray | NMR
Validation Server
BioSync Beamlines/Facilities
Related Tools

Search — Hide
Advanced Search
Latest Release
New Structure Papers
Sequence Search
Chemical Components
Unreleased Entries
Browse Database
Histograms

Tools — Hide
Download: Entries | Ligands
Compare Structures
FTP Services
File Formats

A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the wwPDB, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

Hide Welcome Message

Featured Molecules — Hide

Structural View of Biology — List View of Archive By: Title | Date | Category

Enzymes

Molecule of the Month:
Rhomboid Protease GlpG
Proteases, enzymes that cut protein chains, come in many shapes and sizes. The most familiar proteases, like trypsin and pepsin, are machines of destruction used to digest proteins in our diet. However, most of the proteases in our cells are used in a more delicate task. They regulate the action of other proteins by making specific cuts in their protein targets.
Full Article...

Protein Structure Initiative Featured System:
Exploring the Secretome of Gut Bacteria
JCSG researchers are studying the secretome of the bacteria that inhabit the human gut, to reveal how we live in harmony with these symbiotic organisms.
Full Article | Archive | PSI Structural Biology Knowledgebase

Latest Structures — Hide
3skr - Crystal structure of the 2'- Deoxyguanosine riboswitch bound to 2'-

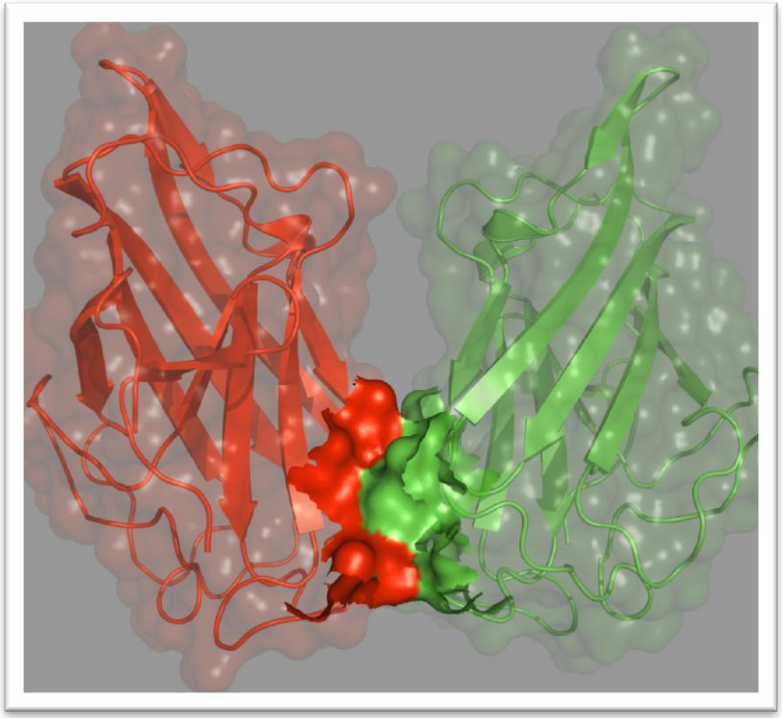
New Features — Hide
Advanced Search: Protein Modifications
Latest features released:
Website Release Archive:

wwPDB News — Hide
PDB40 Symposium
October 28 - 30, 2011
Cold Spring Harbor Laboratory
2011-08-11
IUCr XXII: wwPDB Q&A session, exhibit booth, presentations, posters and more
- Due Aug 1: PDB40 Travel Award Applications for Early Career Scientists and Students
- Full wwPDB News
- Statement on Retraction of PDB Entries

RCSB PDB News — Hide
Weekly | Quarterly | Yearly
2011-08-16
wwPDB Events at IUCr (August 22 -30, 2011)

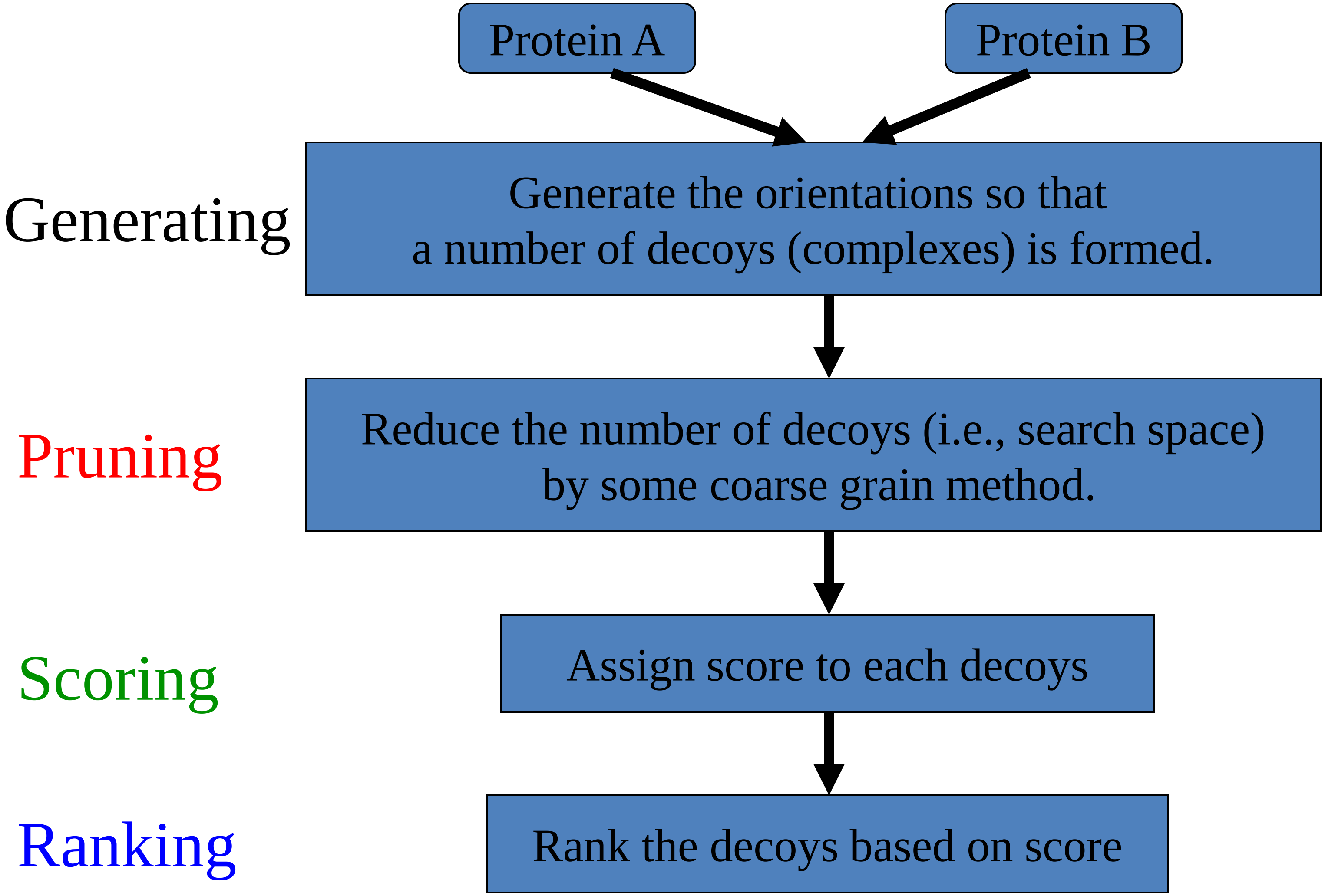
Calderwood M A et al. (2007) *PNAS* **104**, 7606-7611.

Protein Interaction



- Identify the interacting surface for two given protein molecules.
    - To model the protein complex structure formed out of two protein molecules – protein-protein docking

Docking Strategy



Protein A

Protein B

Generating

Generate the orientations so that a number of decoys (complexes) is formed.

Pruning

Reduce the number of decoys (i.e., search space) by some coarse grain method.

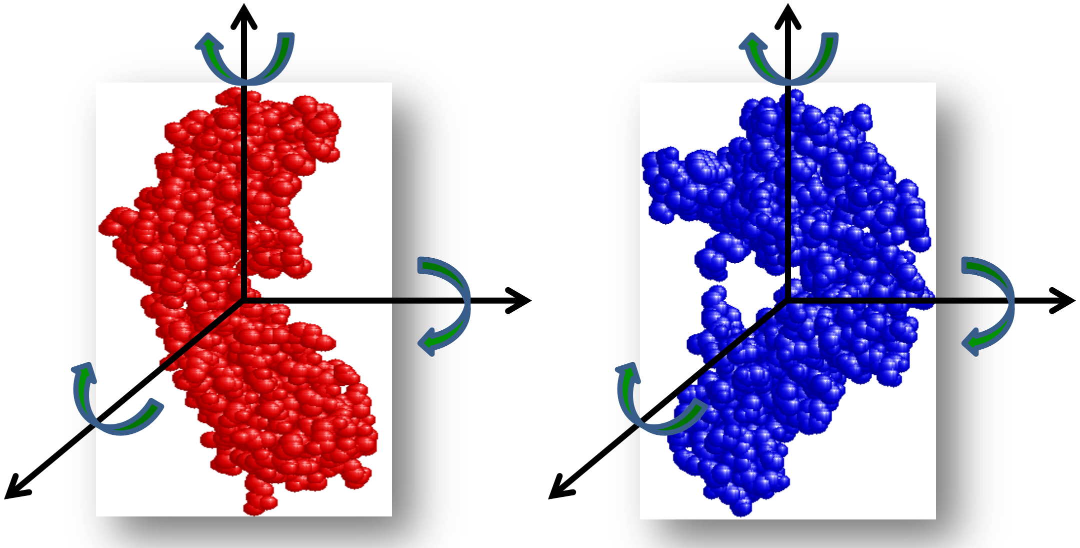
Scoring

Assign score to each decoys

Ranking

Rank the decoys based on score

Generation methods



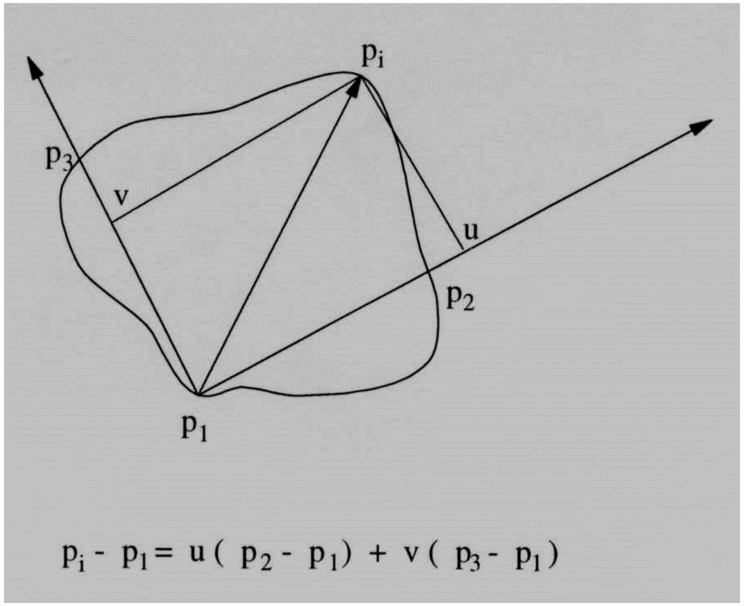
- Tagline – “Higher the decoys; better the possibility of having a hit”
- How many is good?
- Move to discrete space

Geometric Hashing

- Models are represented in a redundant affine invariant way and stored in a table (off-line).
- Hashing is used for organizing and searching the table.

Affine Invariants

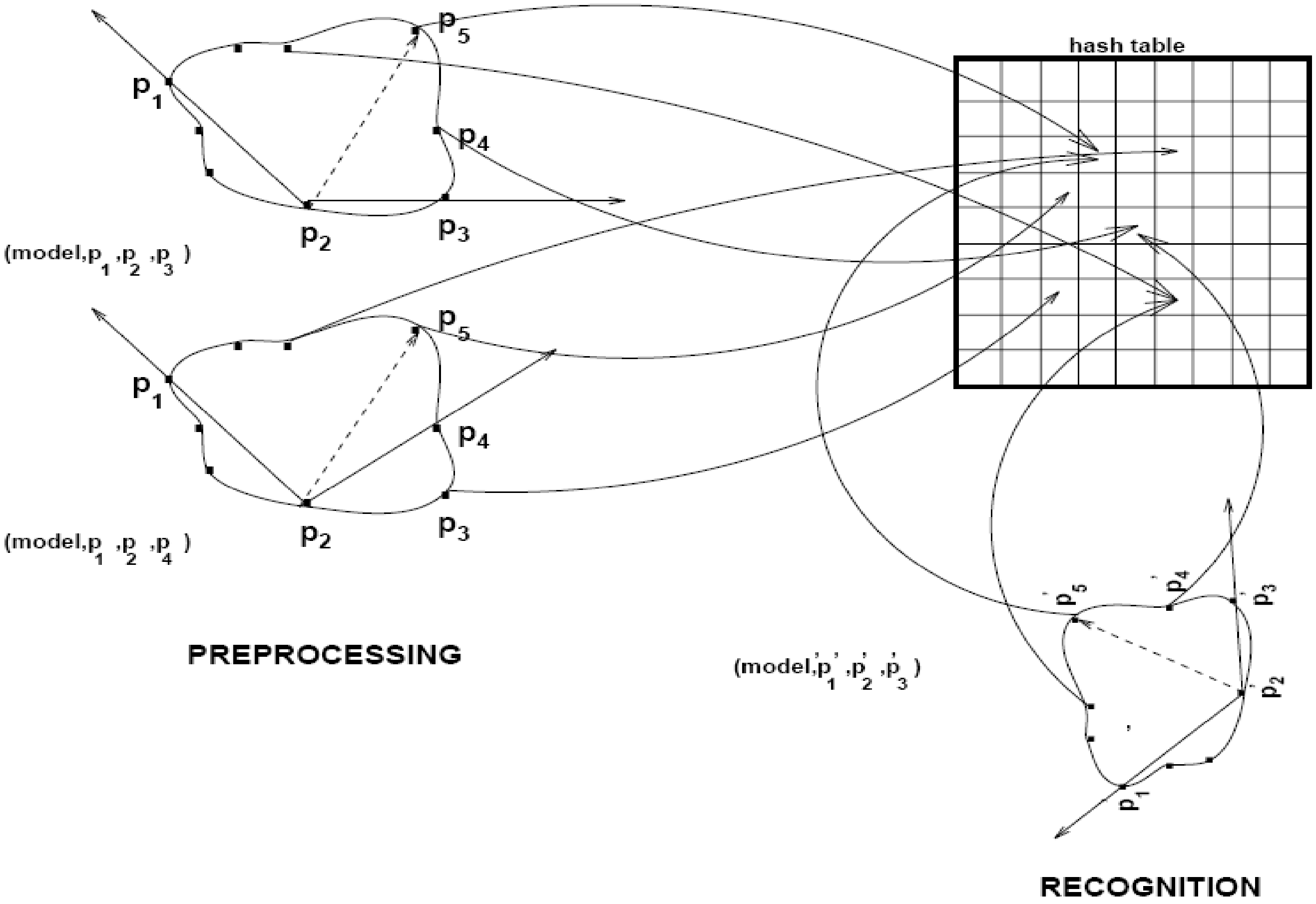
- Each triplet of non-collinear model points forms a basis of a coordinate system that is invariant under affine transformations.
- Represent model points in an affine invariant way by rewriting them in terms of this coordinate system.



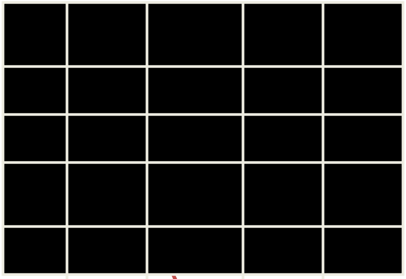
$$p_i - p_1 = u\,(p_2 - p_1) + v\,(p_3 - p_1)$$

(u,v) are affine invariant!

Preprocessing and Recognition



Preprocessing Step



For each model do:

(1) Extract model's point features.

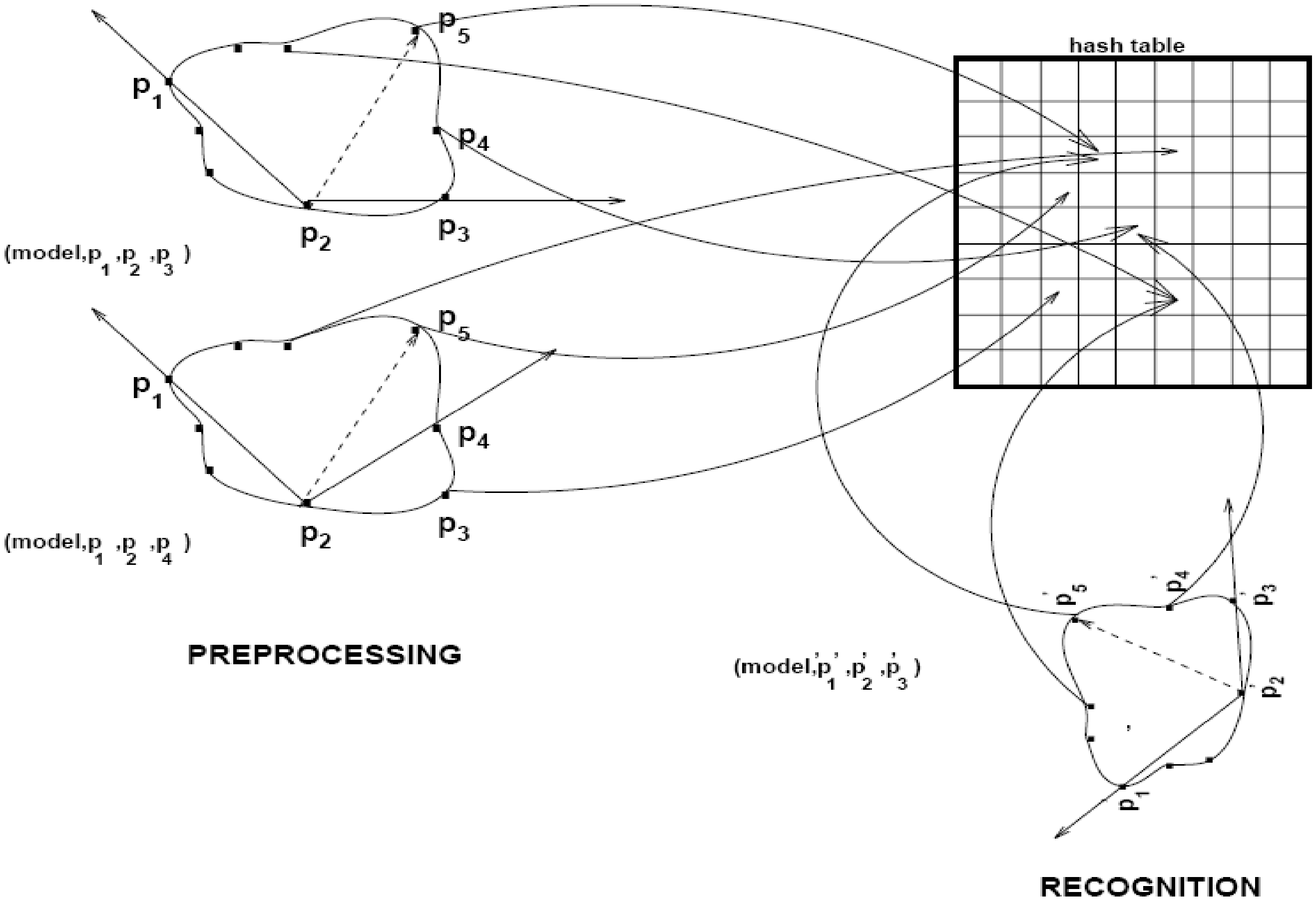
(2) For each ordered set of three, non-collinear, points (p_1, p_2, p_3)

(a) Compute the coordinates **(u,v)** of the remaining features in the coordinate frame defined by the model basis **(p_1, p_2, p_3)**

(b) After a proper quantization, use the computed coordinates **(u,v)** as an index to a two dimensional hash table, and record in the corresponding hash table bin the information **(model, (p_1, p_2, p_3))**

Hash Function: **$h(Q(u), Q(v))$** →

Preprocessing and Recognition



Recognition Step

(1) Extract the image point features

(2) Choose an arbitrary ordered pair **(p'_1, p'_2, p'_3)**

(3) Compute the coordinates (u',v'), of the remaining feature points in the coordinate frame defined by the image basis **(p'_1, p'_2, p'_3)**

(4) After quantization, use the computed coordinates as an index to the hash table. For every entry **(model, (p_1, p_2, p_3))** found in the corresponding bin, cast a vote.

Recognition Step (cont'd)

(5) Histogram all the hash table entries that received one or more votes. Determine those entries that received more than a certain number of votes -- each such entry corresponds to a potential match (hypothesis generation).

(6) For each potential match, consider all the model-image feature pairs which voted for a particular entry, and recover the affine transformation A that results in the best least-squares match between all the corresponding feature points.

Recognition Step (cont’d)

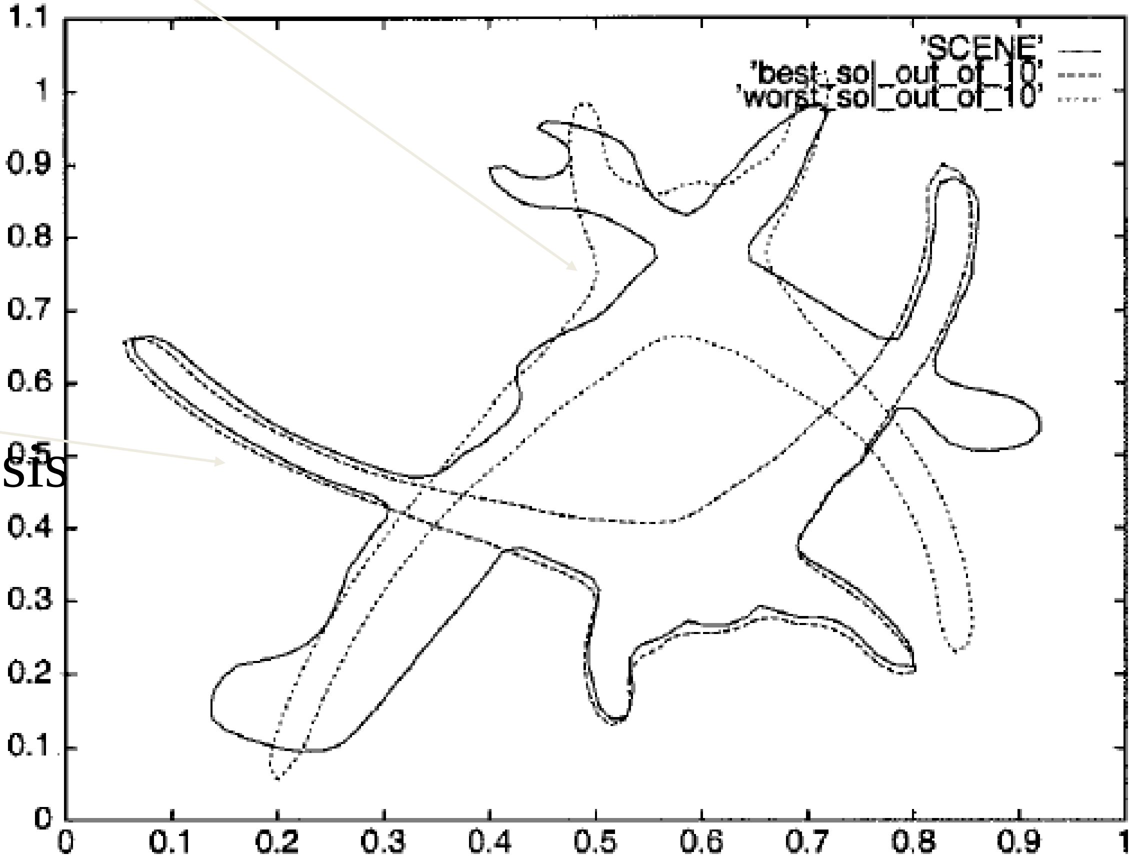
(7) Map the model onto the image using the computed transform and compare the model edges with the image edges (verification step).

(8) If the verification fails for all the models computed in step (5), go back to step (2) and repeat the procedure using a different image basis.

Recognition Example

Bad hypothesis

Good hypothesis



Complexity

- Preprocessing Step:

$O(Mm^4)$

- Recognition Step:

worst case: $O(i^4Mm^4)$

(M: #models, m: #model points, i: #scene points)

Geometric Hashing

- Pro:
  - Faster
- Con:
  - Storage requirement is very high and increases with the increase in object points.
  - Proper identification of object points are crucial for the success.