Recognition

- Three big strategies currently
  - Use geometric reasoning (not widely used now)
  - Match distinctive local patterns with a classifier
  - Find groups of local patterns with discriminative relational properties

Recognition

- Problems
  - detection; localization; kinematics; counting
- Matching
  - Is this a pattern of a fixed class?
    - face detection
  - To what class does this pattern belong?
    - finding faces, animals, motorcycles, etc.
  - Is this pool of patterns consistent with this object?
  - Primary issues:
    - local image representation
    - spatial representation
    - efficiency

Detection

- What pictures contain a giraffe?
- Experimental protocol
  - apply detector to images known to contain/lack object, count
- Relatively easy to get performance figures
  - one doesn’t need to check the giraffe has been put in the right place
  - but they may be meaningless or unreliable
  - in many test sets, objects and backgrounds are strongly correlated
- One should compare performance to baseline
  - e.g. SVM’s on colour histograms; etc.
- Published performance figures are suspect
  - detection rates are implausibly high
  - datasets seldom baselined

Localization

- Where should I shoot to hit the giraffe?
- Experimental protocol unclear
  - how does one score partially correct localization?
  - errors are meaningful only wrt spatial model
- Experiments tricky on a respectable scale
  - but one or two images used to be common
- More difficult criterion to do well at than detection
  - can detect without localizing (detection marginalizes out configuration)
- Few published performance figures

Kinematics and counting

- Kinematics
  - What is the giraffe's configuration?
    - Experimental protocol thoroughly unclear
      - what is a partial success?
      - what does one count?
      - how?
    - Not much known except for human tracking cases
- Counting
  - how many giraffes are there?
    - Experimental protocol easy in principle
    - Obviously, very difficult to do without localization
      - appears to be difficult even with models that can localize
      - we should be able to count things we haven't seen before
        - one of many links between segmentation and recognition
    - No current system can count anything significant satisfactorily

More uncertain technologies

- Relational reasoning
  - Currently
    - Objects are composed of parts; find the parts; are the relations right?
  - Perhaps
    - How are objects distributed in space?
    - Which objects are made of the same stuff?
- Knowledge building
  - Shop around mixed collections to obtain world knowledge
    - building object models; a face dictionary; etc.
- Generalization
  - Map knowledge across kinds of object
    - This <animal> won't bite; this <animal> is scary and about to pounce
  - Requires
    - identifying "kind" (significant component is visual)
    - knowing what can be mapped, and where (mysterious)

Recognition by Hypothesize and Test

- General idea
  - Hypothesize object identity and pose
  - Recover camera (widely known as backprojection)
  - Render object in camera
  - Compare to image
- Issues
  - where do the hypotheses come from?
  - How do we compare to image (verification)?
- Simplest approach
  - Construct a correspondence for all object features to every correctly sized subset of image points
  - These are the hypotheses
  - Expensive search, which is also redundant.

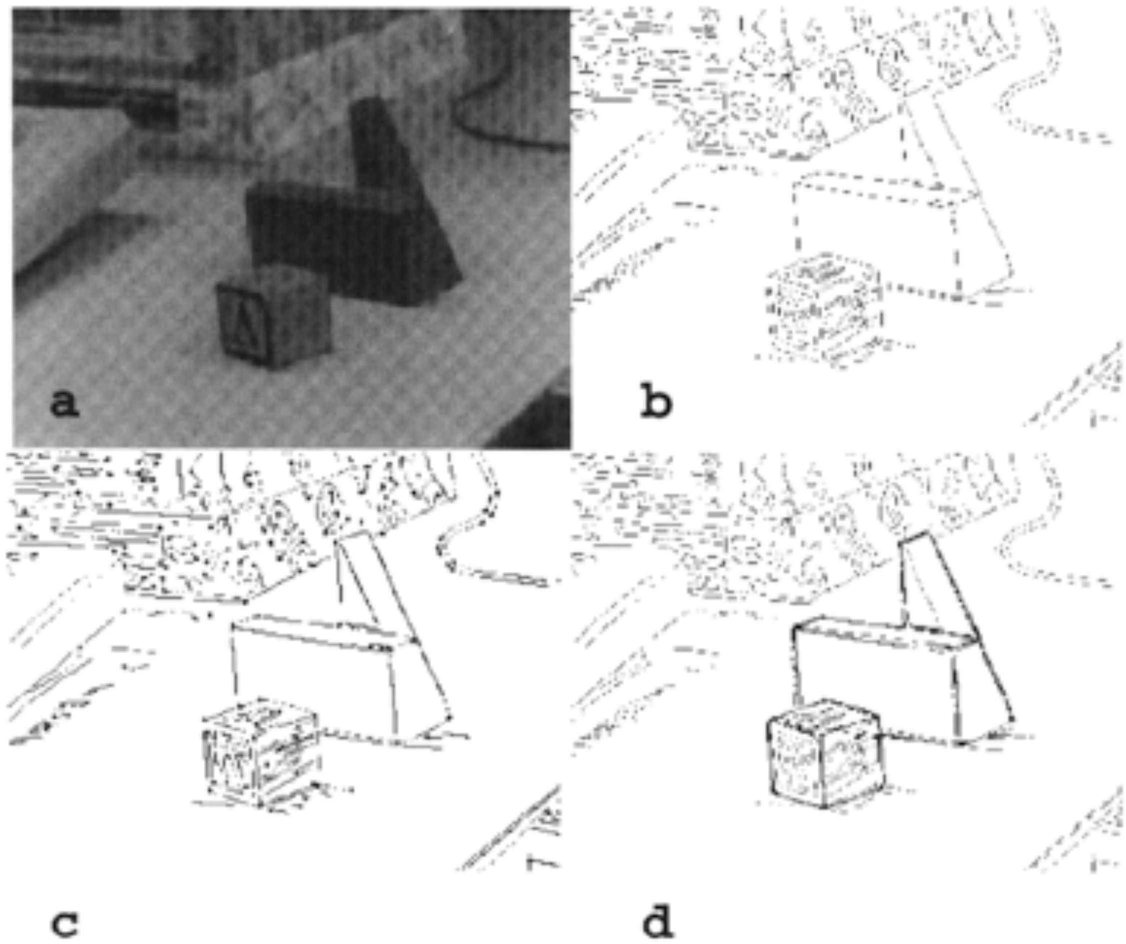
What are the features?

- They have to project like points
  - Lines
  - Conics
  - Other fitted curves
  - Regions (particularly the center of a region, etc.)

Pose consistency

- A small number of correspondences yields a camera
- Strategy:
  - Generate hypotheses using small numbers of correspondences (e.g. triples of points for a calibrated perspective camera, etc., etc.)
  - Backproject and verify
    - Notice that the main issue here is camera calibration
    - Appropriate groups are “frame groups”

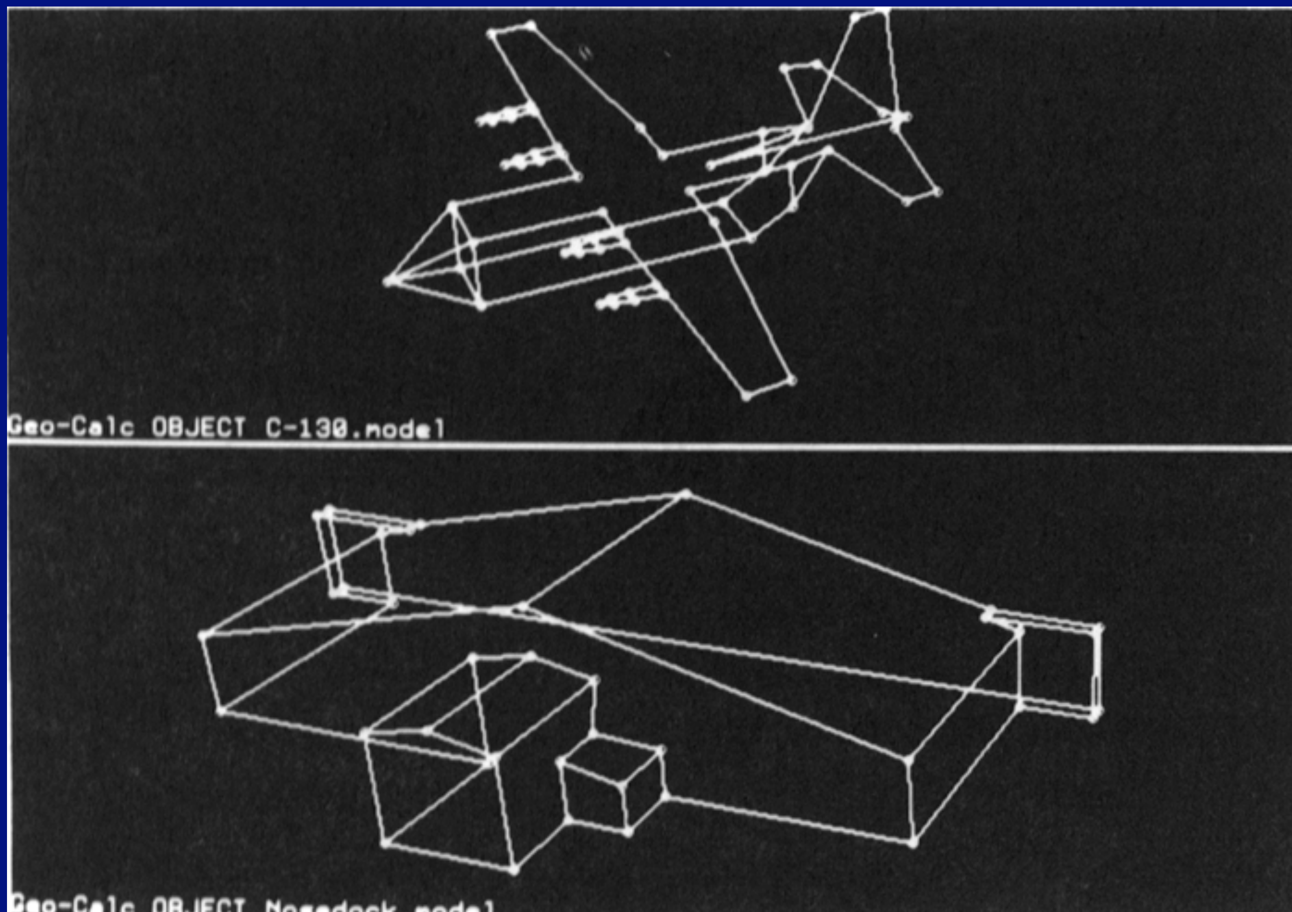
```
For all object frame groups O
  For all image frame groups F
    For all correspondences C between
      elements of F and elements
      of O

      Use F, C and O to infer the missing parameters
      in a camera model

      Use the camera model estimate to render the object

      If the rendering conforms to the image,
        the object is present
    end
  end
end
```

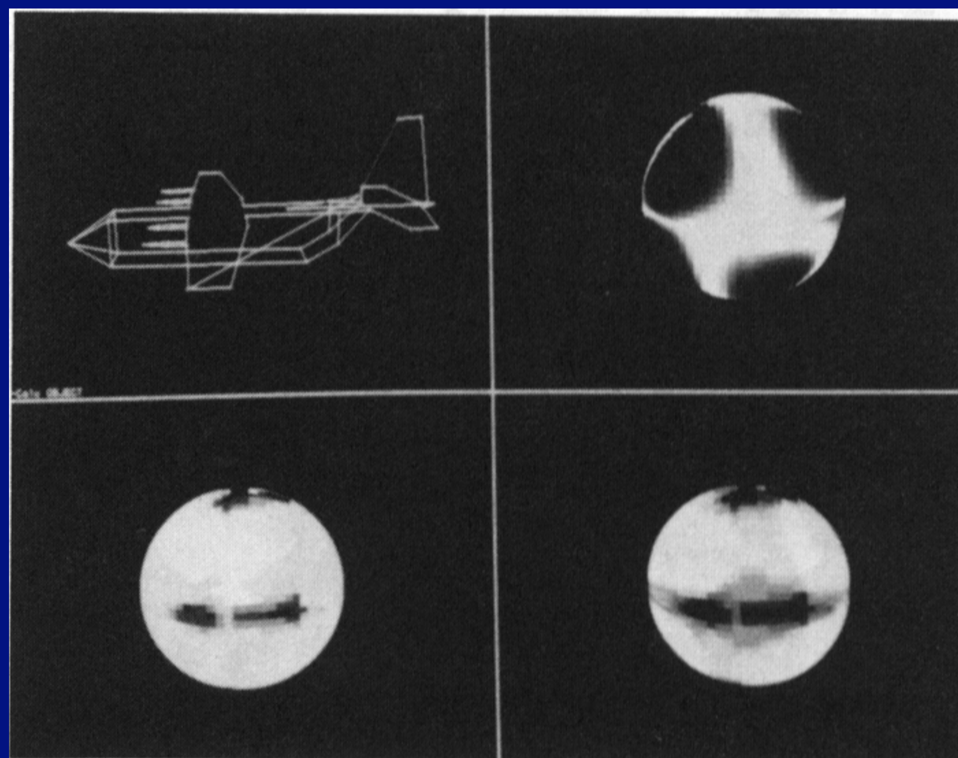

a
b
c
d

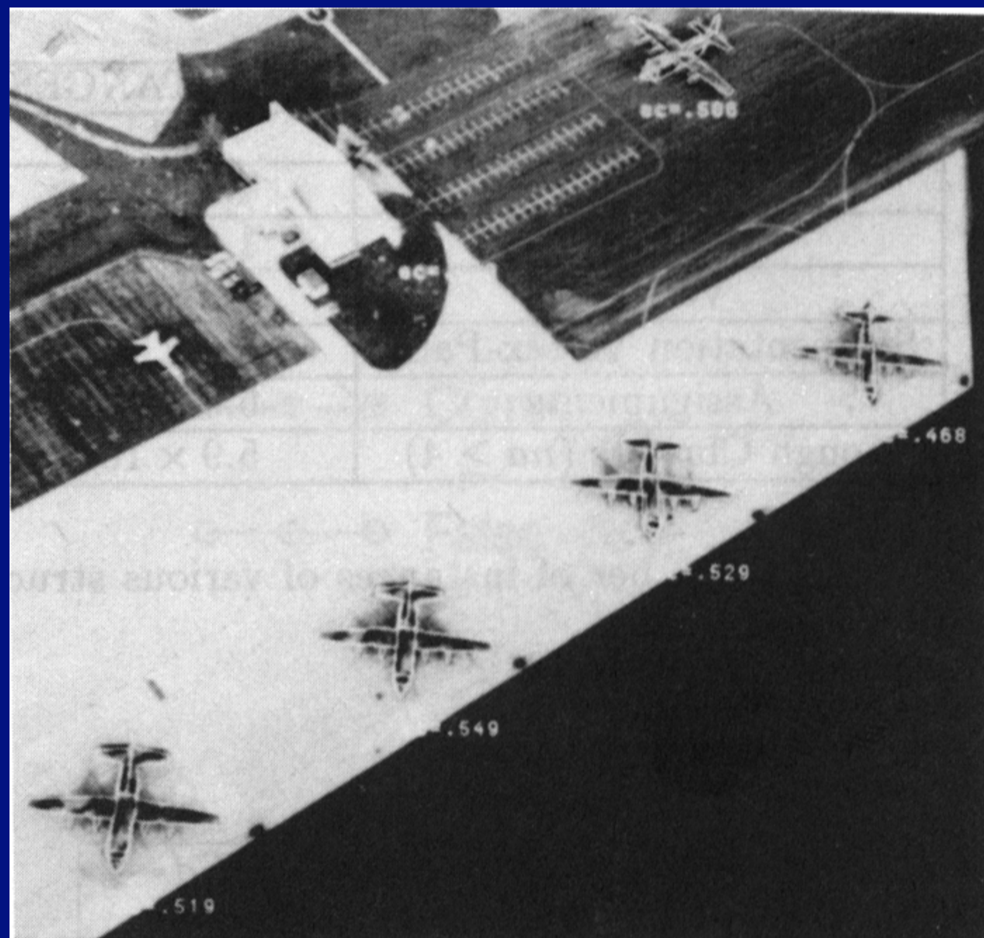
Voting on Pose

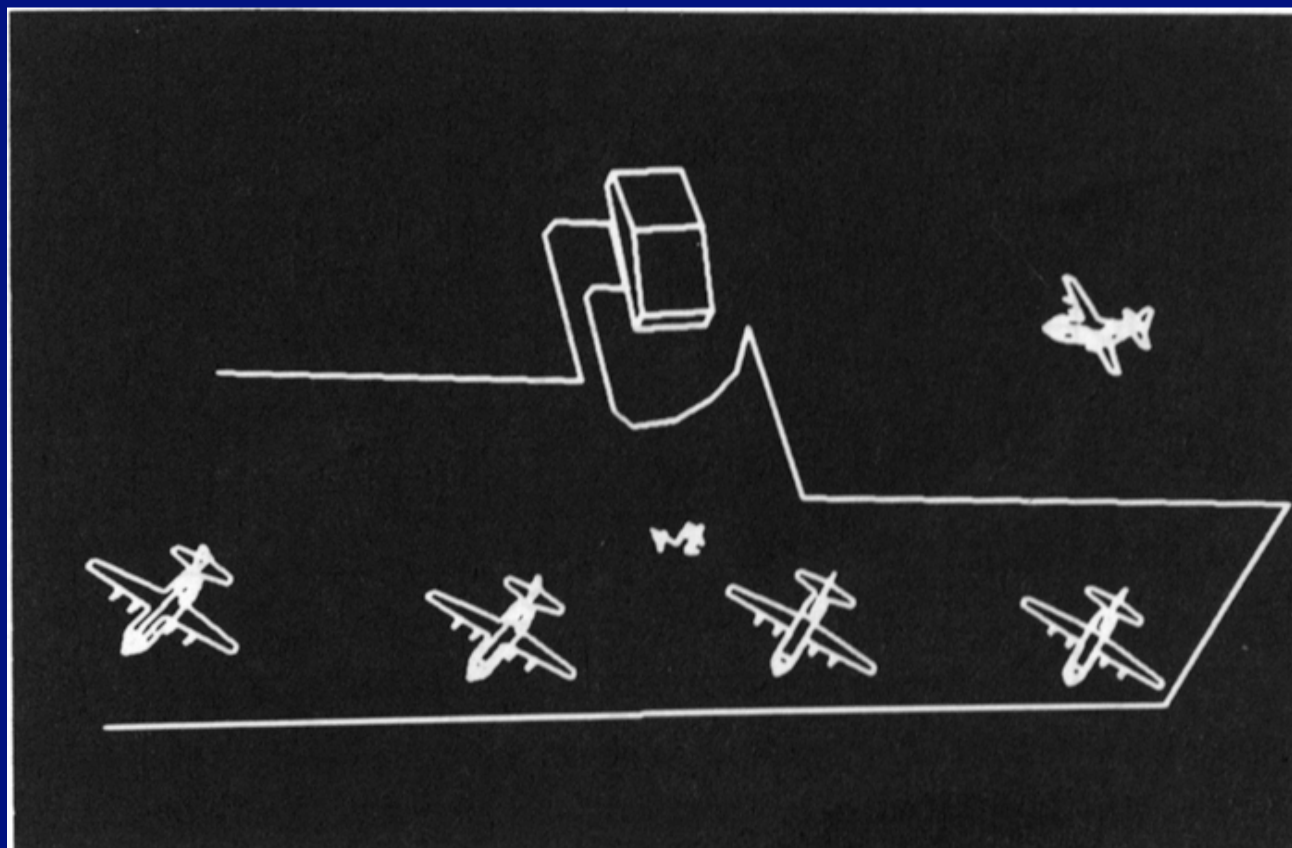
- Each model leads to many correct sets of correspondences, each of which has the same pose
  - Vote on pose, in an accumulator array
  - This is a hough transform, with all it's issues.

```
For all objects O
  For all object frame groups F(O)
    For all image frame groups F(I)
      For all correspondences C between
        elements of F(I) and elements
        of F(O)

        Use F(I), F(O) and C to infer object pose P(O)

        Add a vote to O's pose space at the bucket
        corresponding to P(O).
      end
    end
  end
end
For all objects O
  For all elements P(O) of O's pose space that have
    enough votes

    Use the P(O) and the
    camera model estimate to render the object

    If the rendering conforms to the image,
    the object is present
  end
end
```


Geo-Calc OBJECT C-130.model
Geo-Calc OBJECT Nosedock.model



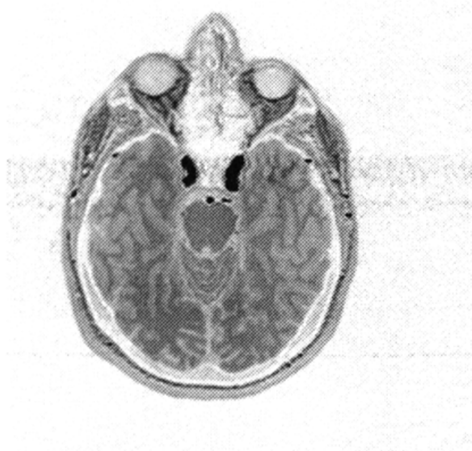

.468
.529
.549
.519

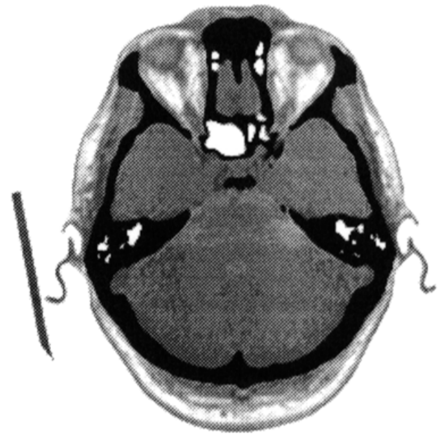


Application: Registering medical images

- To remove only affected tissue
- To minimize damage by operation planning
- To reduce number of operations by planning surgery

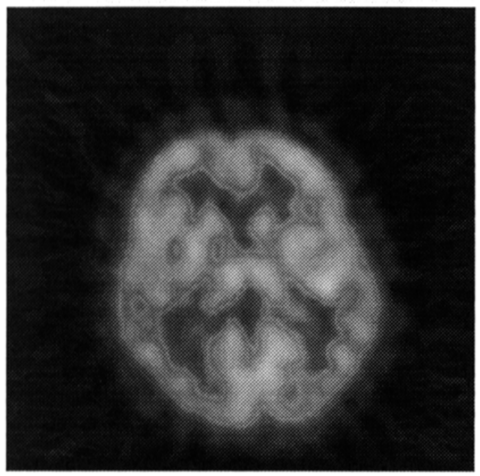
MRI

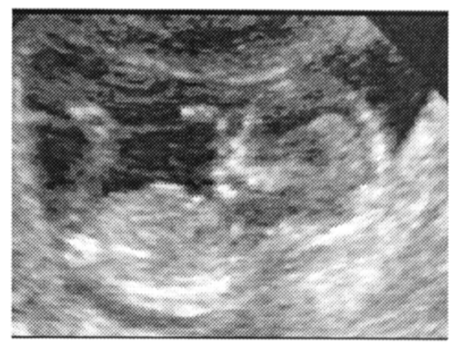




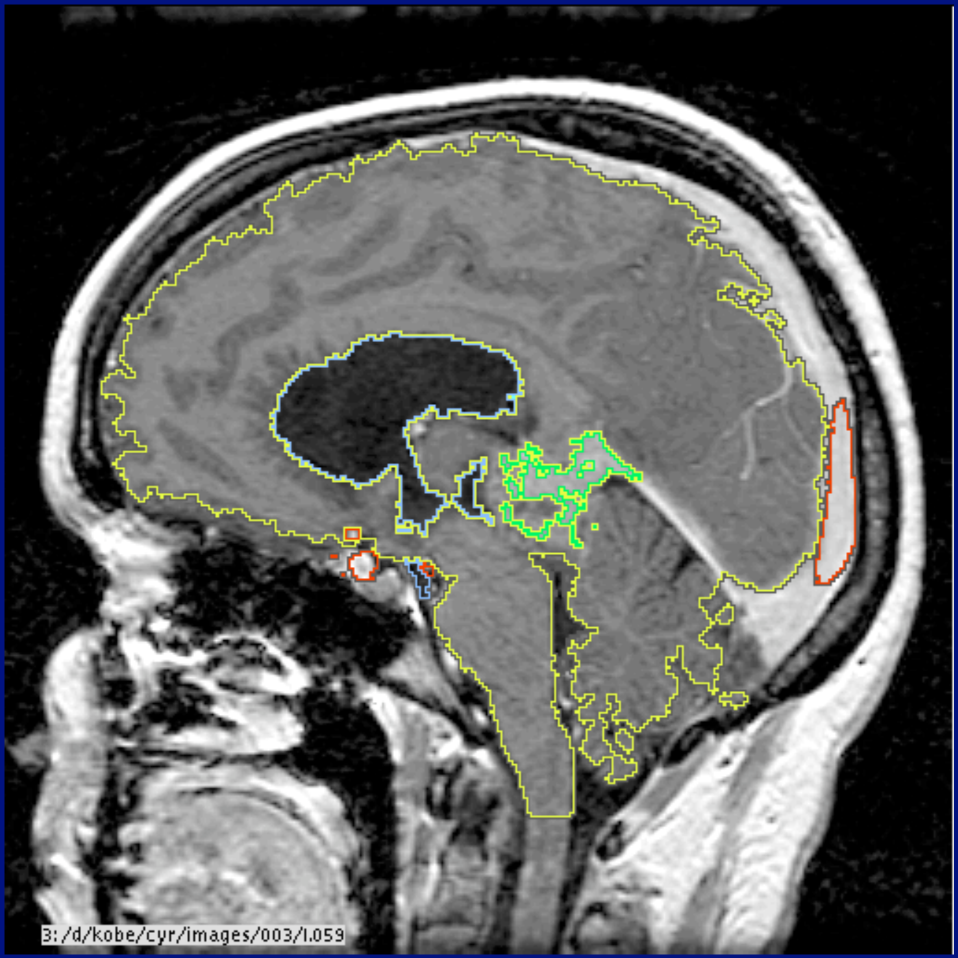
CTI

NMI





USI



Images courtesy of Eric Grimson