Progress on: Variable Viewpoint Reality Image Database

Paul Viola & Eric Grimson

Jeremy DeBonet, Aparna Lakshmiratan, William Wells, Kinh Tieu,Dan Snow, Owen Ozier, John Winn, Mike Ross

MIT Artificial Intelligence Lab

Overview of Presentation

- Variable Viewpoint Reality
  - Overview
  - Progress at MIT
- Image Database Retrieval
  - Overview
  - Progress
- http://www.ai.mit.edu/projects/NTTCollaboration

VVR: Motivating Scenario



- *Construct a system that will allow each/every user to observe any viewpoint of a sporting event.*
- *Provide high level commentary/statistics*
  - *Analyze plays*

For example …

Computed using a single view…
some steps by hand

VVR Spectator Environment

- Build an exciting, fun, high-profile system
  - Sports: Soccer, Hockey, Tennis, Basketball
  - Drama, Dance, Ballet
- Leverage MIT technology in:
  - Vision/Video Analysis
    - Tracking, Calibration, Action Recognition
    - Image/Video Databases
  - Graphics
- Build a system that provides data available nowhere else…
  - Record/Study Human movements and actions
  - Motion Capture / Motion Generation

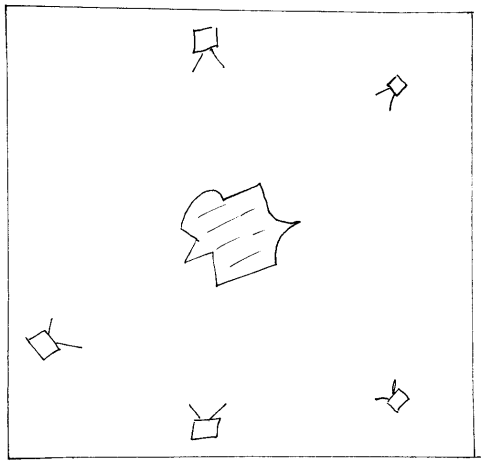
Window of Opportunity

- 20-50 cameras in a stadium
  - Soon there will be many more
- US HDTV is digital
  - Flexible, very high bandwidth digital transmissions
- Future Televisions will be Computers
  - Plenty of extra computation available
  - 3D Graphics hardware will be integrated
- Economics of sports
  - Dollar investments by broadcasters is huge (Billions)
- Computation is getting cheaper

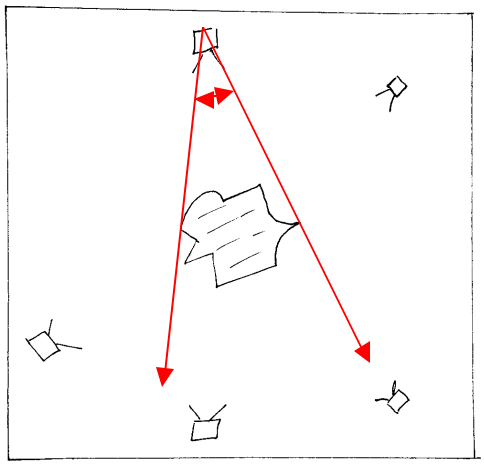
Progress at MIT

- Simple intersection of silhouettes (Visual Hull)
  - Efficient but limited
- Tomographic reconstruction
  - Based on medical reconstruction
- Probabilistic Voxel Analysis (Poxels)
  - Handles occlusion & transparency
- Parametric Human Forms

Visual Hull in 2D

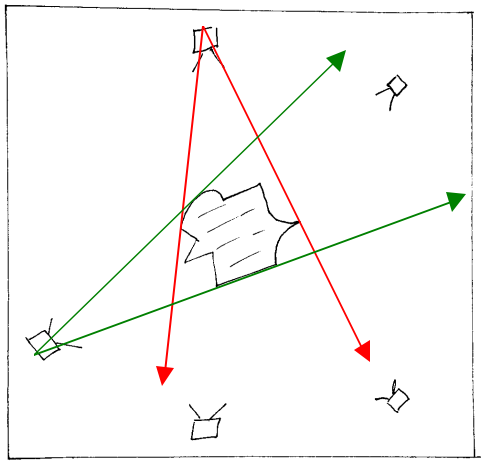


Visual Hull: Segment



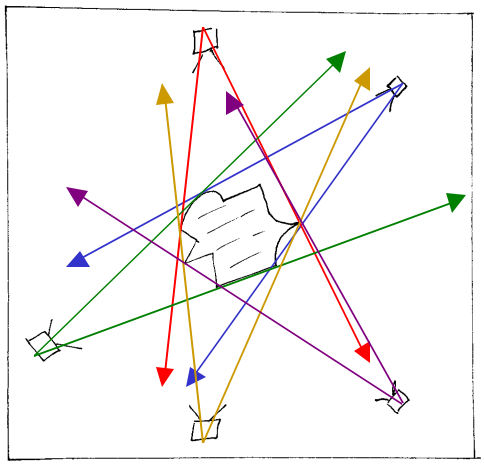
Viola 1999
MIT AI Laboratory

Visual Hull: Segment



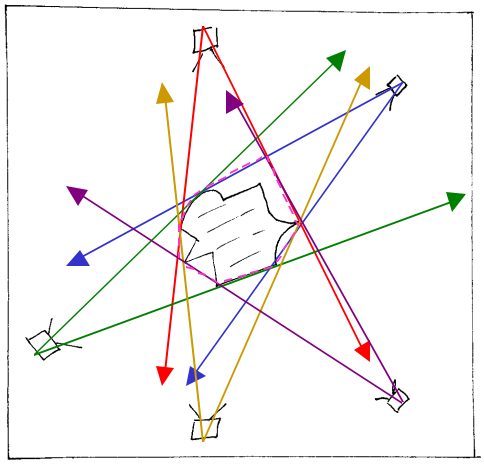
Viola 1999
MIT AI Laboratory

Visual Hull: Segment



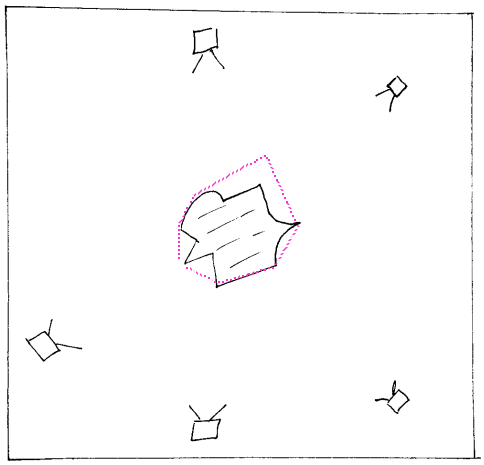
Viola 1999
MIT AI Laboratory

Visual Hull: Intersection



Viola 1999
MIT AI Laboratory

Idea in 2D: Visual Hull

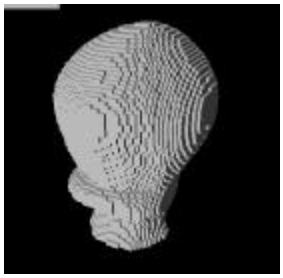


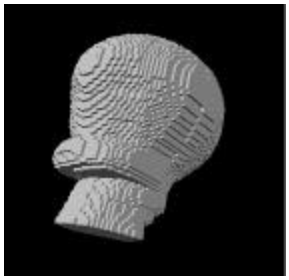
Real Data: Tweety



- Data acquired on a turntable
  - 180 views are available… not all are used.

Intersection of Frusta

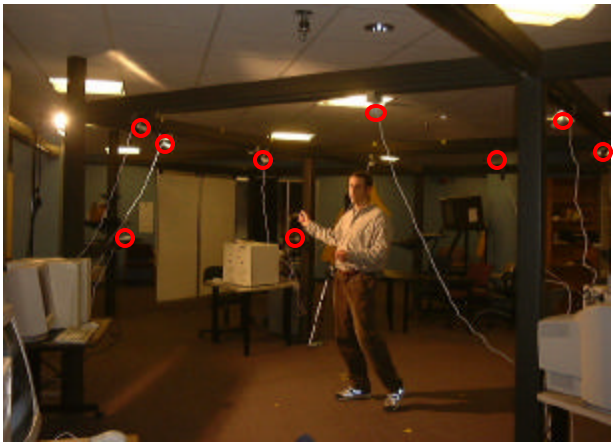




- Intersection of 18 frusta
  - Computations are very fast
    - perhaps real-time

Viola 1999
MIT AI Laboratory

New Apparatus

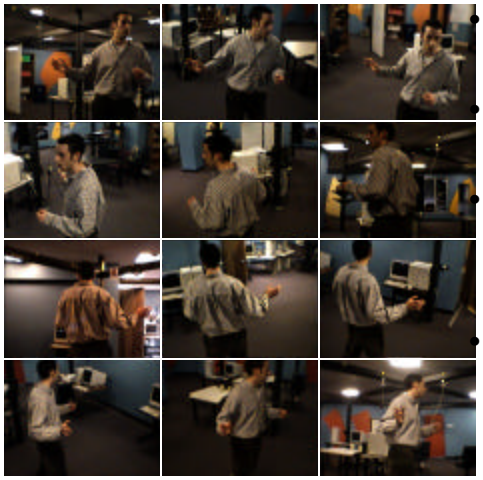


Twelve cameras, computers, digitizers

Parallel software for acquisition/processing

Viola 1999
MIT AI Laboratory

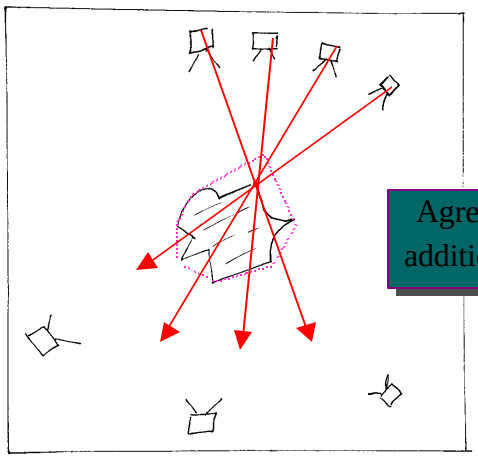
Current System



- Real-time image acquisition
- Silhouettes computed in parallel
- Silhouettes sent to a central machine
  - 15 per second
- Real-time Intersection and Visual Hull
  - In progress

Viola 1999
MIT AI Laboratory

Visual Hull is very coarse …

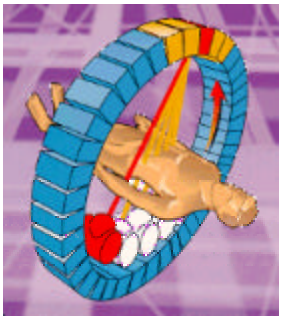


Agreement provides additional information

Viola 1999
MIT AI Laboratory

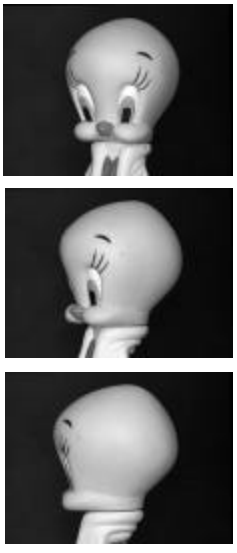
Tomographic Reconstruction

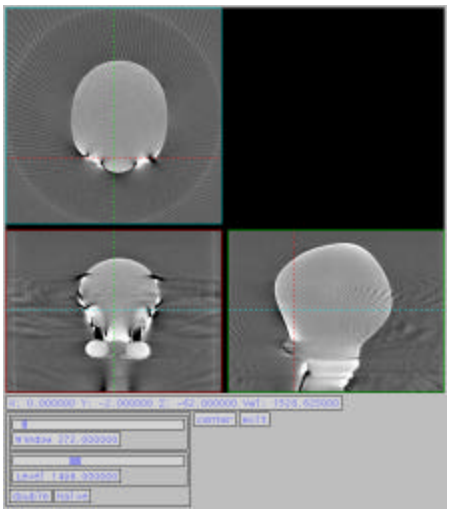
- Motivated by medical imaging
  - CT - Computed Tomography
  - Measurements are line integrals in a volume
  - Reconstruction is by back-projection & *deconvolution*



Viola 1999
MIT AI Laboratory

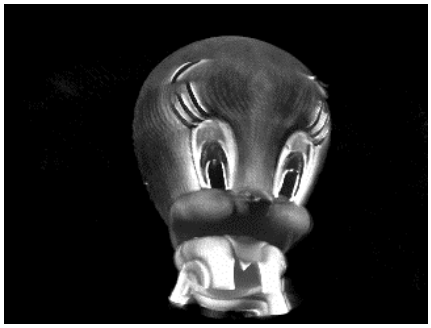
Back-projection of image intensities





Viola 1999
MIT AI Laboratory

Volume Render...

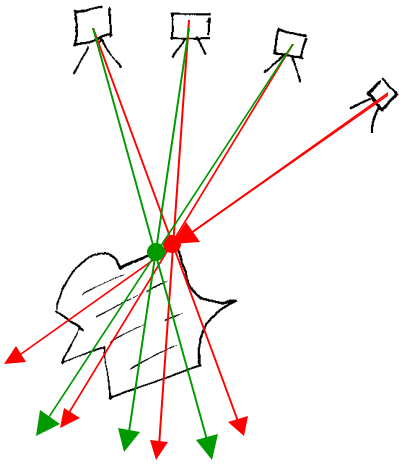


- Captures shape very well
- Intensities are not perfect

Poxels: An improvement to tomography

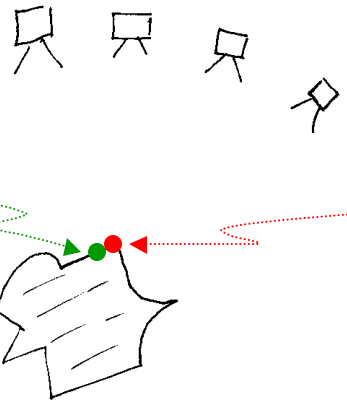
- Tomography confuses color with transparency
  - Does not model occlusion...
- The Probabilistic Voxel Approach: Poxel
  - Estimates both color and transparency
  - Models occlusion
  - Much better results
    - Though slower
  - Work submitted to ICCV 99

Occlusion causes disagreement



Viola 1999
MIT AI Laboratory

Initial agreement is not enough…

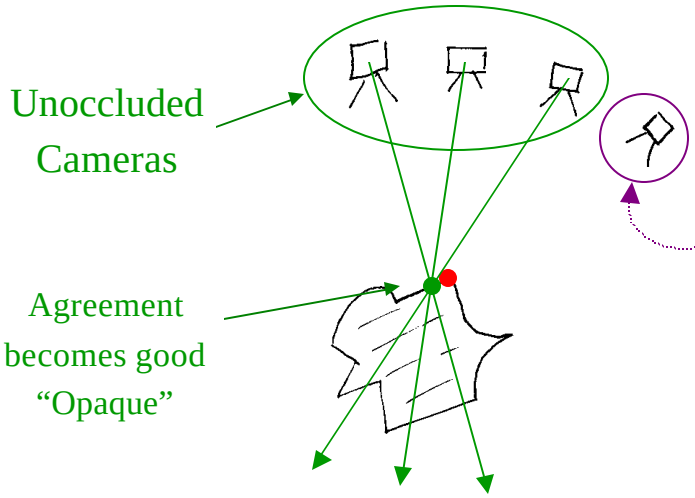


Agreement is poor

Agreement is high *"Opaque"*

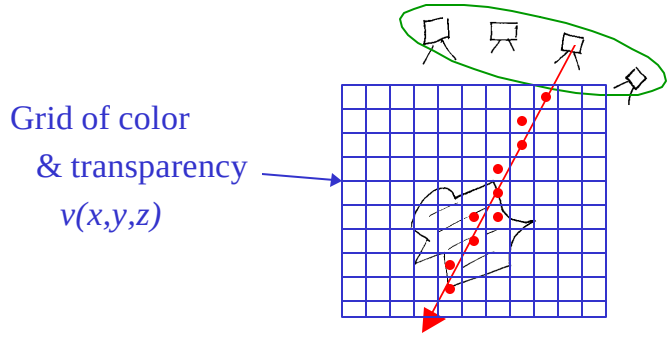
Viola 1999
MIT AI Laboratory

Second pass uses information about occlusion



Viola 1999
MIT AI Laboratory

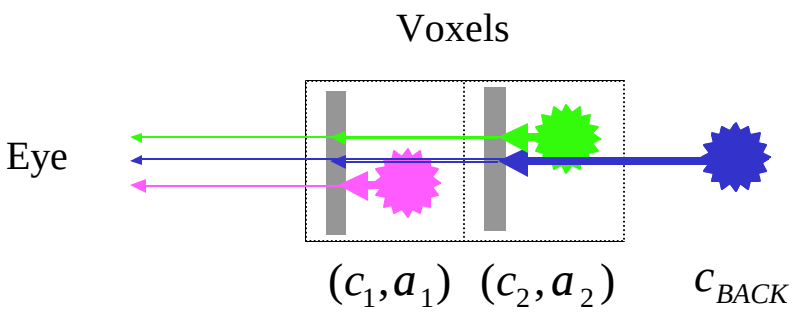
Poxels Algorithm: Definitions



$c(x,y,z)$	*colors*
$a(x,y,z)$	*transparancy*
$v(x,y,z)$	*voxels (color+trans)*
$I_k(u,v)$	*images*
$l_k(u,v,d)=(x,y,z)$	*ray castingimagek*

Viola 1999
MIT AI Laboratory

Poxels: Model of Transparency

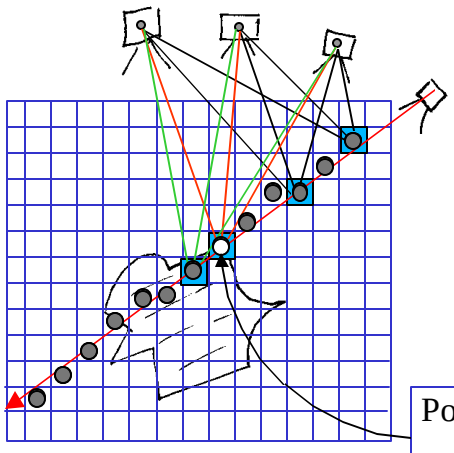


$$\text{color} = a_1\left(c_1 + a_2\left(c_2 + c_{BACK}\right)\right)$$

$$\hat{I}_k(u,v) = v(l_k(u,v,1)) \wedge v(l_k(u,v,2)) \wedge \Lambda \wedge v(l_k(u,v,d))$$

Viola 1999
MIT AI Laboratory

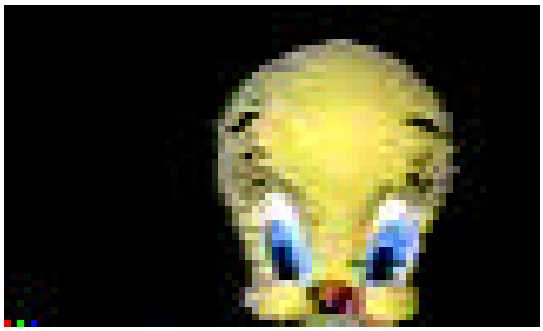
Poxels Algorithm: Agreement (Step 1)



Point of highest agreement

Viola 1999
MIT AI Laboratory

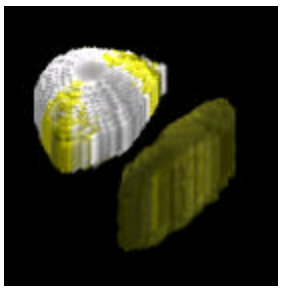
Results…



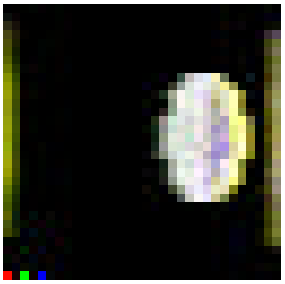
Rendering of reconstructed shape.

From ICCV paper...

Input Image



Reconstructed Volume



… additional results

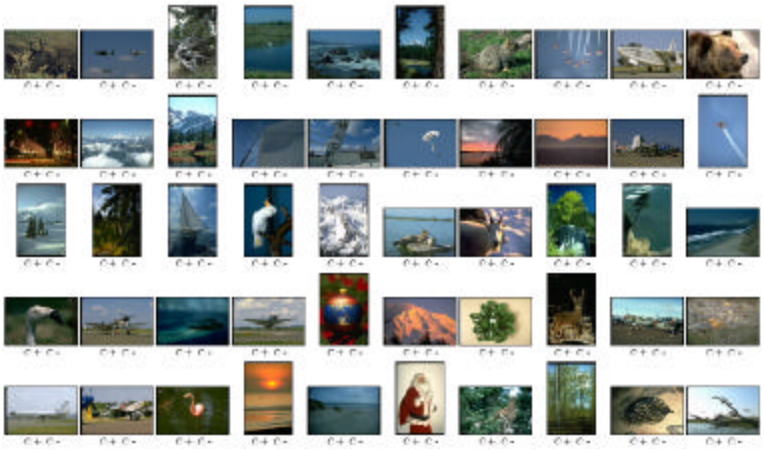
Viola 1999
MIT AI Laboratory

Image Databases: Motivating Scenario

- Image Databases are proliferating
  - The Web
  - Commercial Image Databases
  - Video Databases
  - Catalog Databases
    - "Find me a bag that looks like a Gucci."
  - Virtual Museums
    - "Find me impressionist portraits."
  - Travel Information
    - "Find me towns with Gothic architecture."
  - Real-estate
    - "Find me a home that is sunny and open."

Viola 1999
MIT AI Laboratory

But, the problem is very hard…



There are a very wide variety of images...

Viola 1999
MIT AI Laboratory

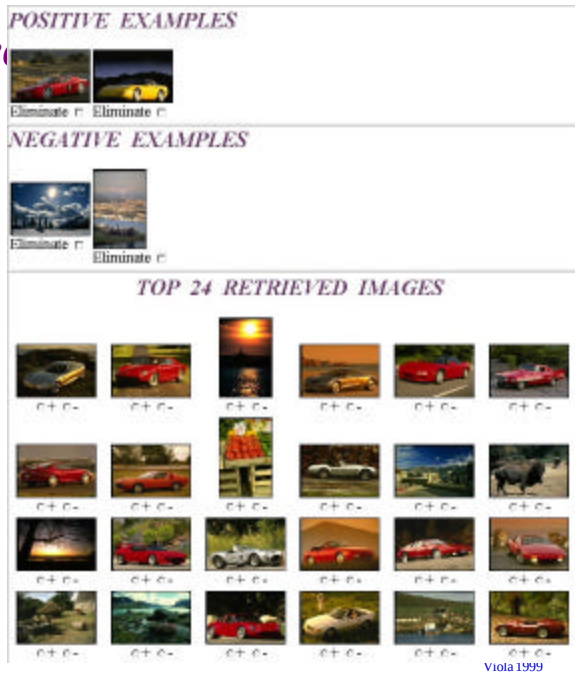
We have made good progress...

Query: “Waterfall Images”

Viola 1999
MIT AI Laboratory

Se

Trained by example



Viola 1999
MIT AI Laboratory

Complex Feature Representation

- Motivated by the Human brain…
  - Infero-temporal cortex computes many thousand selective features
  - Features are selective yet insensitive to unimportant variations
  - Every object/image has some but not all of these features
- Retrieval involves matching the most salient features

Viola 1999
MIT AI Laboratory

Image Database Retrieval
NTT: Visit
1/7/99

Overview of IDB Meeting

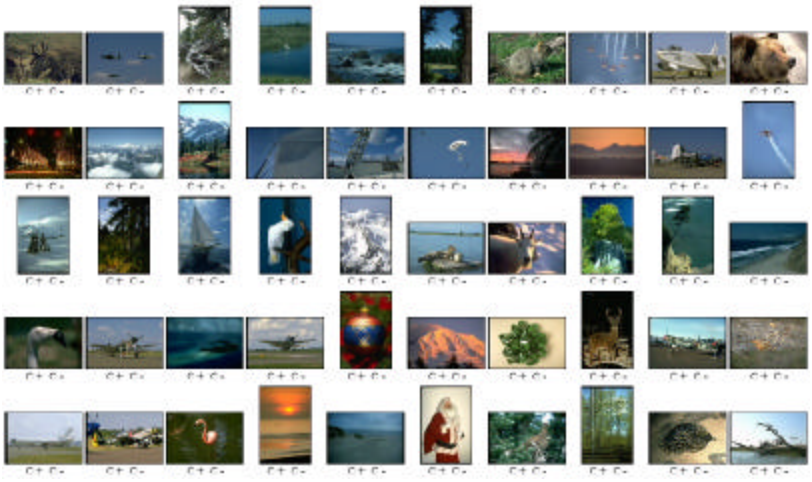
- Motivation from MIT ...
- Discuss current and related work
  - Flexible Templates
  - Complex Features
  - Demonstrations
- Related NTT Efforts
- Discussion of collaboration
- Future work
- Dinner

Motivating Scenario

- Image Databases are proliferating
  - The Web
  - Commercial Image Databases
  - Video Databases
  - Catalog Databases
    - "Find me a bag that looks like a Gucci."
  - Virtual Museums
    - "Find me impressionist portraits."
  - Travel Information
    - "Find me towns with Gothic architecture."
  - Real-estate
    - "Find me a home that is sunny and open."

Viola 1999
MIT AI Laboratory

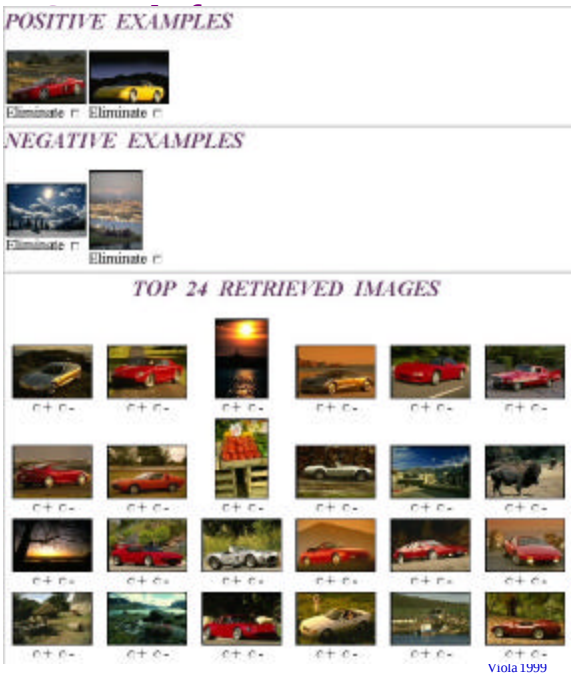
There is a very wide variety of images...



Viola 1999
MIT AI Laboratory

Search for images containing waterfalls?

Viola 1999
MIT AI Laboratory



Viola 1999
MIT AI Laboratory

What makes IDB hard?

- Finding the right features
  - Insensitive to movement of components
  - Sensitive to critical properties
- Focussing attention
  - Not everything matters
- Generalization based on class
  - Given two images
    - Small black dog & Large white dog
    - (Don't have much in common…)
  - Return other dogs

Viola 1999
MIT AI Laboratory

Overview of IDB Meeting

- Motivation from MIT ...
- Discuss current and related work
  - Flexible Templates
  - Complex Features
  - Demonstrations
- Related NTT Efforts
- Discussion of collaboration
- Future work
- Dinner

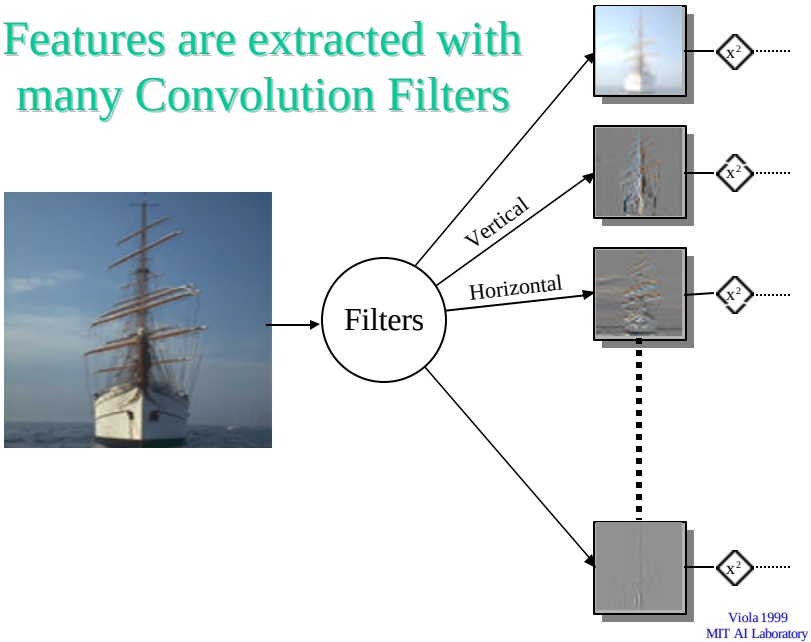
Viola 1999
MIT AI Laboratory

Complex Feature Representation

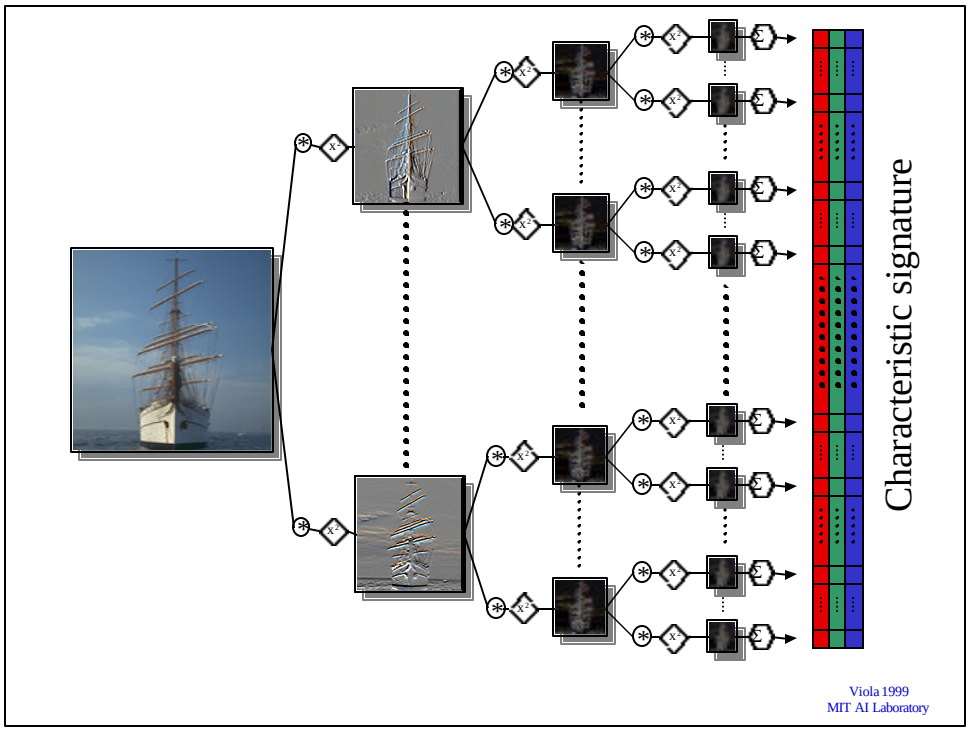
- Motivated by the Human brain…
  - Infero-temporal cortex computes many thousand selective features
  - Features are selective yet insensitive to unimportant variations
  - Every object/image has some but not all of these features
- Retrieval involves matching the most salient features

Viola 1999
MIT AI Laboratory

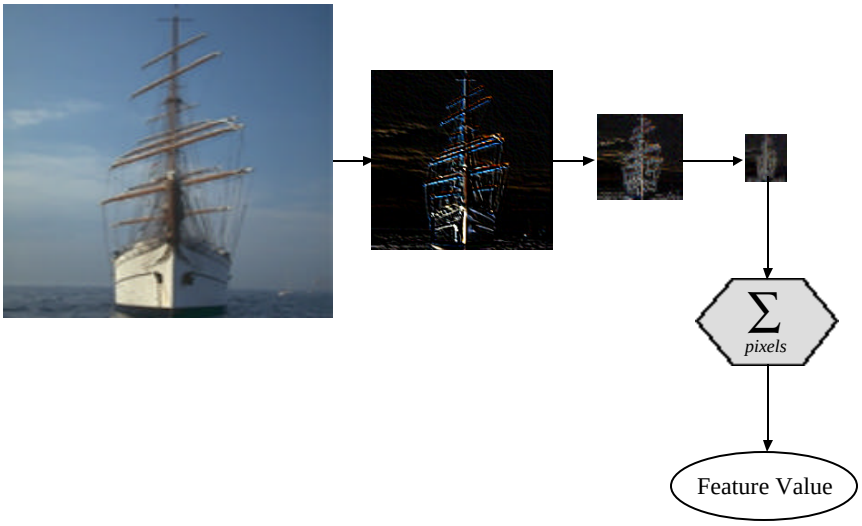
Features are extracted with many Convolution Filters



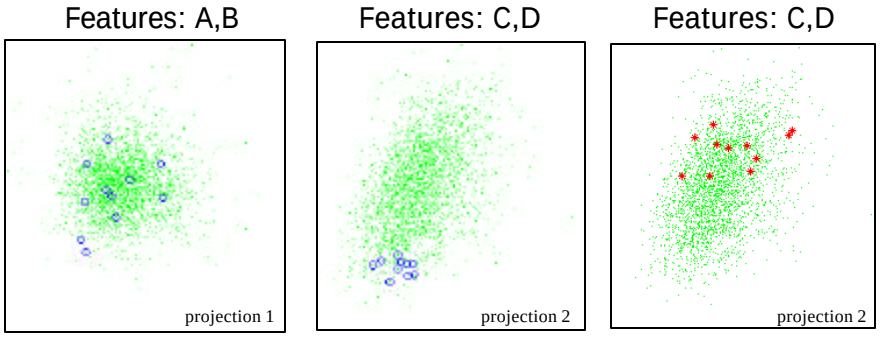
Viola 1999
MIT AI Laboratory



Resolution is reduced at each step…



Not every feature is useful for a query



Poor Features for query 1

Good Features for query 1

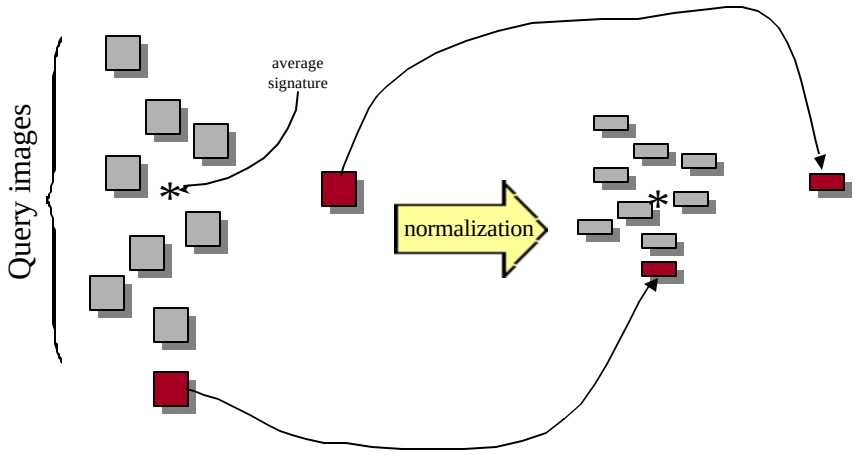
Poor Features for query 2

Query 1 (variation of object location)

Query 2 (variation of lighting)

Viola 1999
MIT AI Laboratory

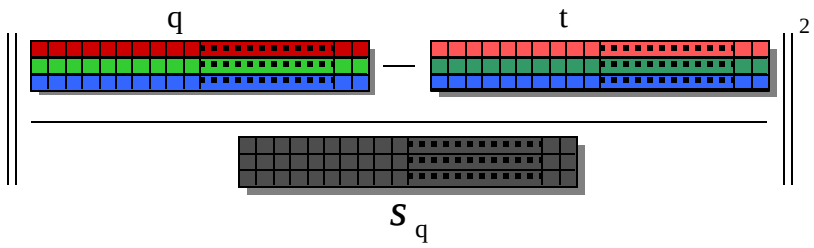
Normalization of Signature Space



•Normalization brings some image closer to the mean

Viola 1999
MIT AI Laboratory

Distance/Similarity Measure



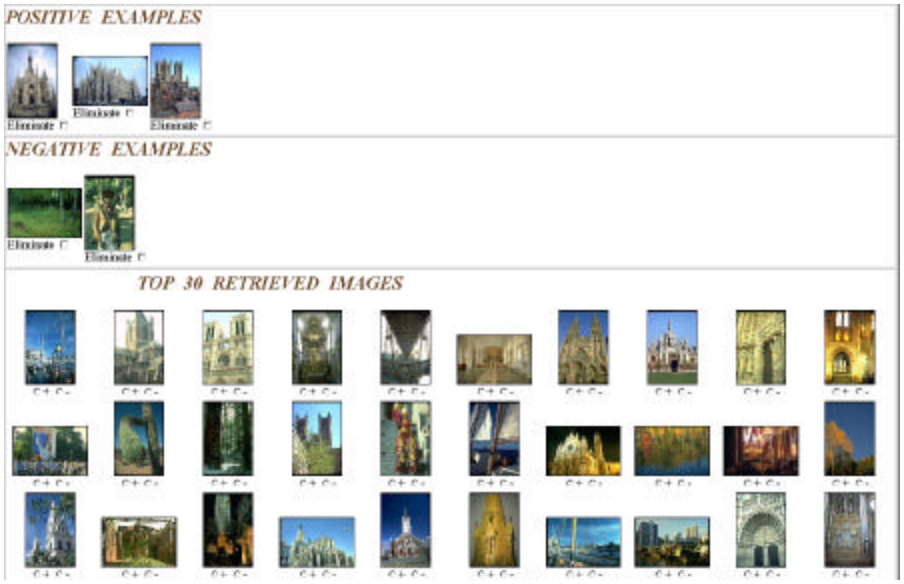
$$\sum_{e=0}^{25^3} \sum_{c=\{r,g,b\}} \left(\frac{\overline{q_{e,c}} - t_{e,c}}{s_{q_{e,c}}} \right)^2$$

Diagonal Mahalanobis Distance

Image Database Progress at MIT

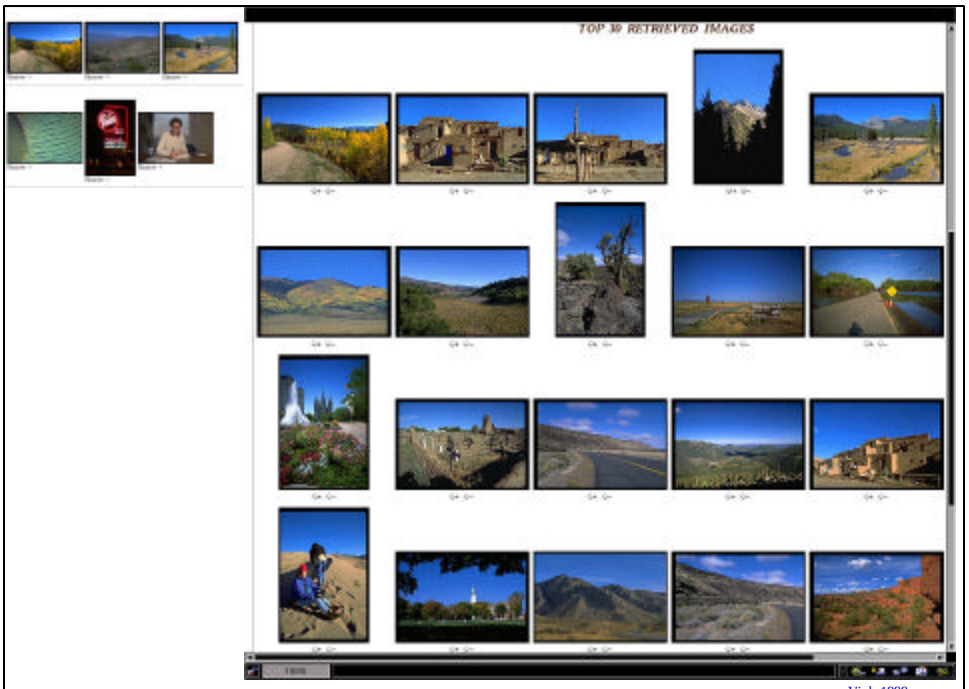
- Better learning algorithms to select features
- Developed a very compact feature representation
  - Fewer features required
  - 2-3 bits per feature

  } factor of 72
- Pre-segmentation of images
  - Better learning
  - More selective queries
- Construction of object models:
  - Faces, people, cars, etc. (ICCV 99)

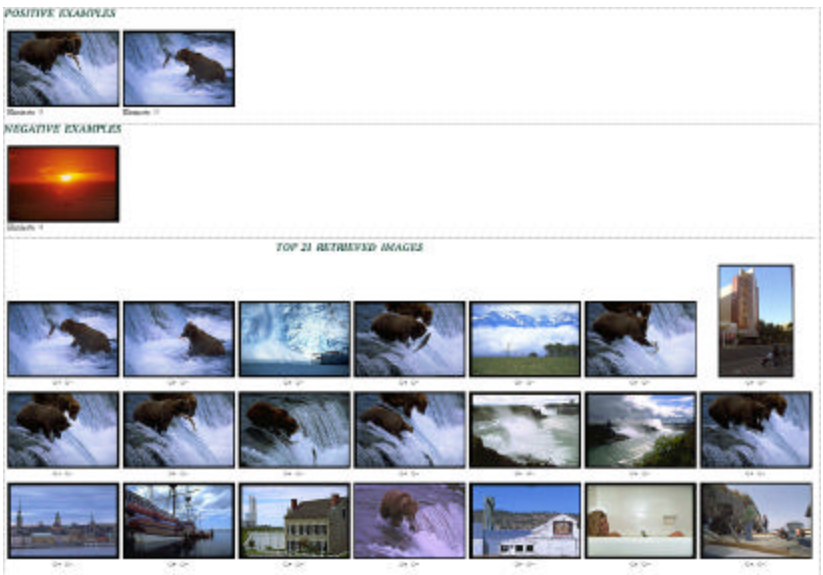


Viola 1999
MIT AI Laboratory

ola 1999
I Laboratory



Viola 1999
MIT AI Laboratory



Viola 1999
MIT AI Laboratory

Conclusions

- Variable Viewpoint Reality
  - Prototypes constructed
  - New approaches
- Image Database Retrieval
  - New more efficient representations
  - Improved performance