Overview

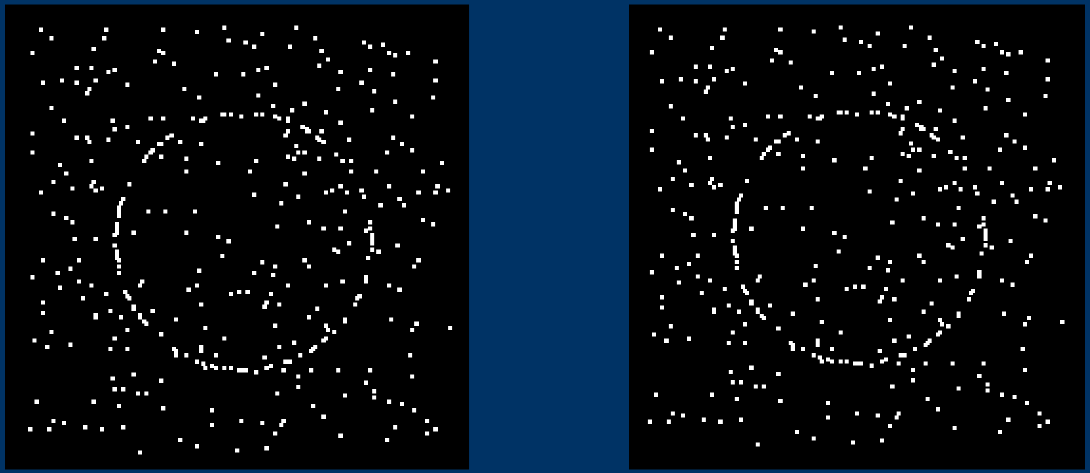
- Related Work
- Tensor Voting in 2-D
- Tensor Voting in 3-D
- Tensor Voting in N-D
- Application to Vision Problems
- Stereo
- Visual Motion
- Binary-Space-Partitioned Images
- 3-D Surface Extraction from Medical Data
- Epipolar Geometry Estimation for Non-static Scenes
- Image Repairing
- Range and 3-D Data Repairing
- Video Repairing
- Luminance Correction
- Conclusions

Visual Motion Analysis

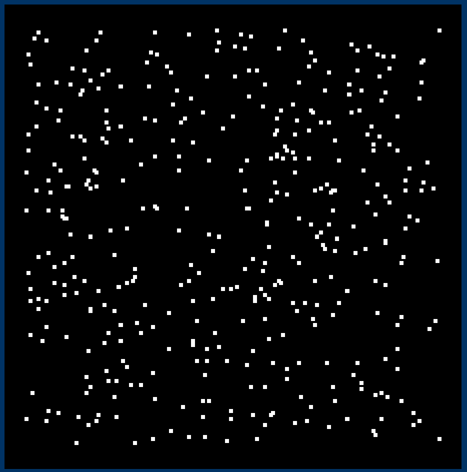
- From motion cues only
- From real images

Monocular vs. Motion Cues

Structure inference possible from one image only…?



…or from motion only ?

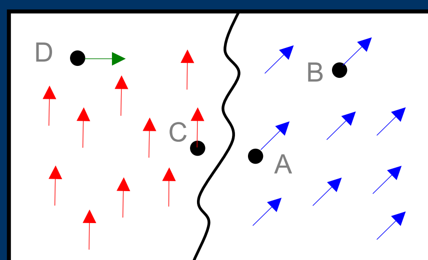


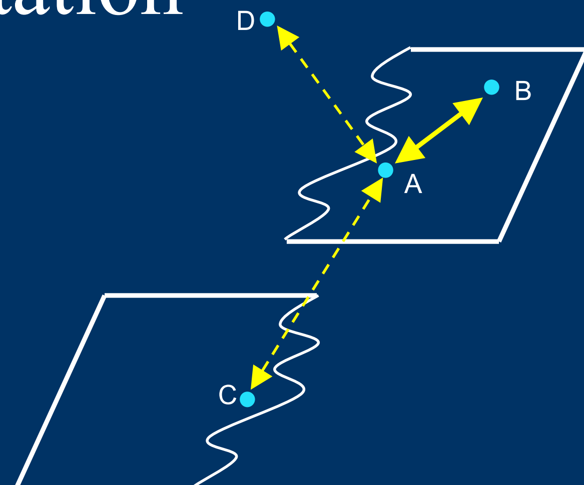
Computational Processes

- Matching
  - Establish token correspondences across images
  - Recover a (possibly sparse and noisy) velocity field
- Motion capture
  - Obtain a dense representation :
  - Dense velocity field
  - Boundaries
  - Regions

4-D Voting Approach

- Layered 4-D representation





- Match: $(x\ y) \rightarrow (x+v_x\ y+v_y)$
- Represent each candidate match as a $(x\ y\ v_x\ v_y)$ point in 4-D
- Motion layers $\leftrightarrow$ smooth surfaces in the 4-D space

Second Order Tensors in 4-D

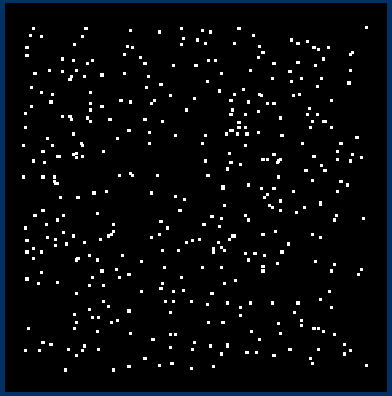
Elementary tensors

Feature	λ_1	λ_2	λ_3	λ_4	e_1	e_2	e_3	e_4	Tensor
Point	1	1	1	1	Any orthonormal basis				Ball
Curve	1	1	1	0	n_1	n_2	n_3	t	C-Plate
Surface	1	1	0	0	n_1	n_2	t_1	t_2	S-Plate
Volume	1	0	0	0	n	t_1	t_2	t_3	Stick

A generic tensor

Feature	Saliency	Normals	Tangents
Point	λ_4	none	none
Curve	$\lambda_3 - \lambda_4$	e_1 e_2 e_3	e_4
Surface	$\lambda_2 - \lambda_3$	e_1 e_2	e_3 e_4
Volume	$\lambda_1 - \lambda_2$	e_1	e_2 e_3 e_4

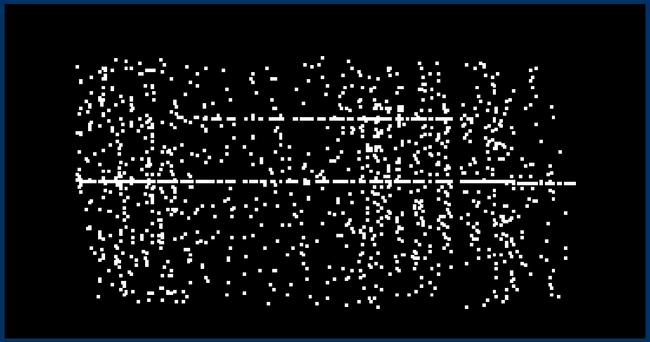
Generating Candidate Matches



Input images:

- sparse identical point tokens
- motion cues only

- Establish a potential match with all tokens in a neighborhood

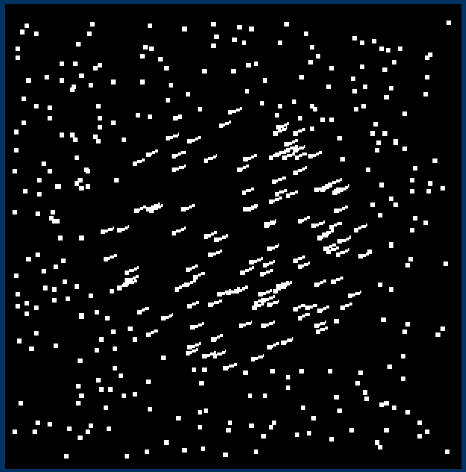


Candidate matches:

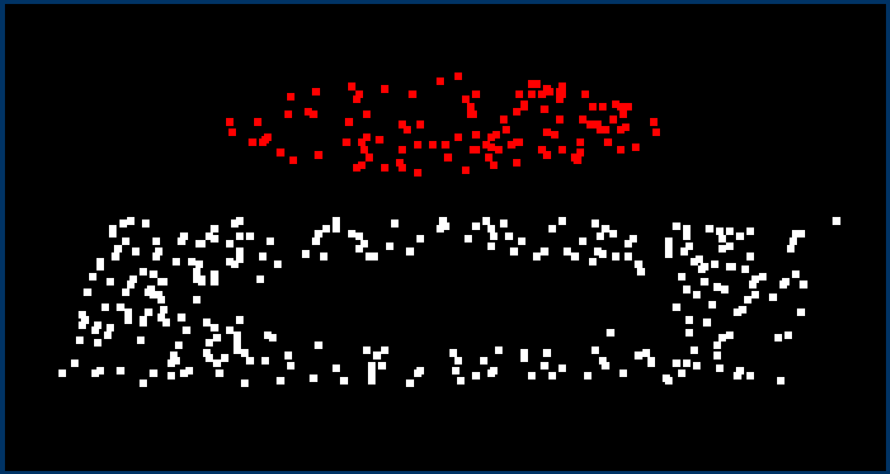
- $(x\ y\ v_x\ v_y)$ points in 4-D

Selection

- Wrong matches appear as outliers, receiving little or no support
- Affinity (support) is expressed by the surface saliency at each token: λ_2 - λ_3

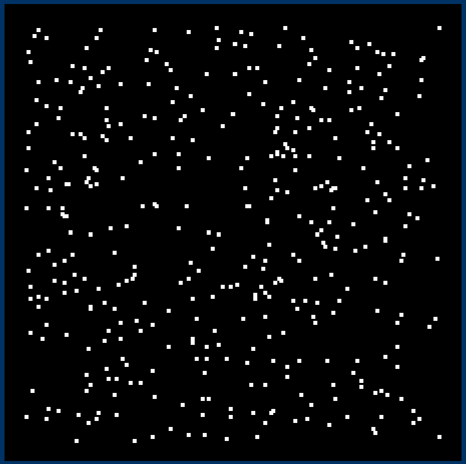


Sparse velocity field



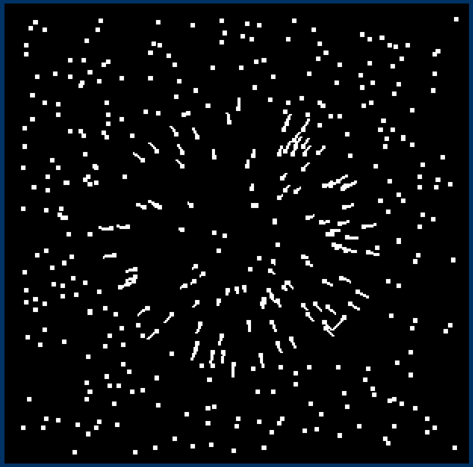
Recovered v_x velocities

Expanding Disk

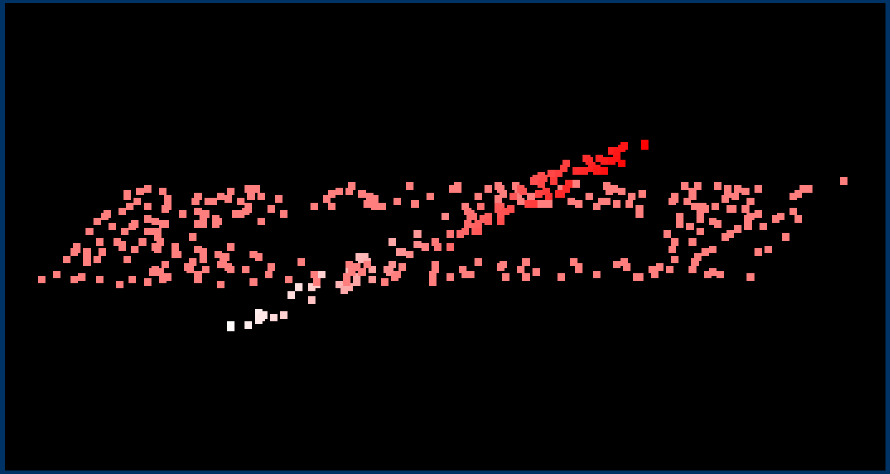


Input

Non-rigid motion

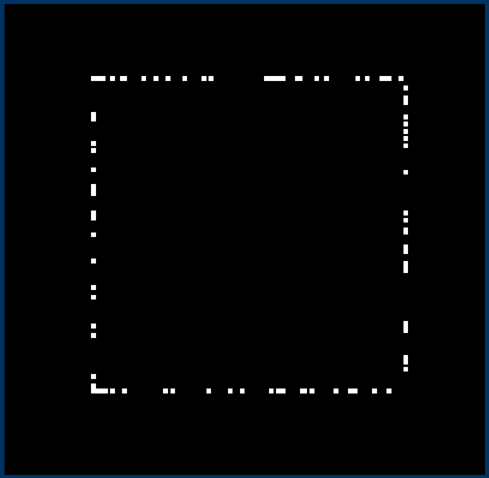


Sparse velocity field



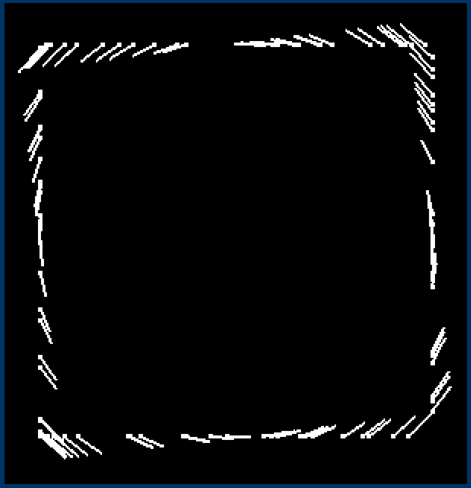
3-D view of recovered v_x velocities

Rotating Square

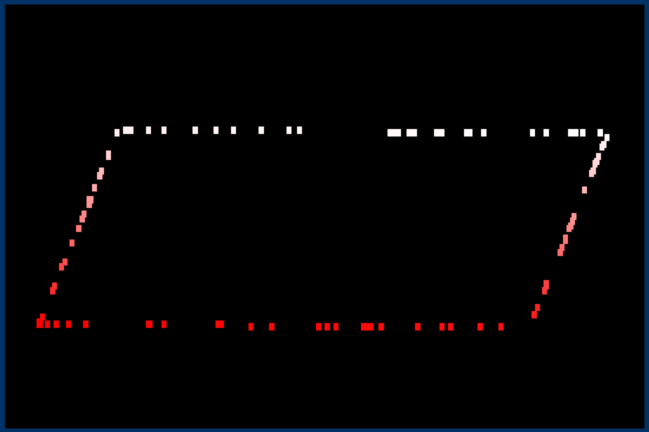


Input

Non-smooth curve



Sparse velocity field

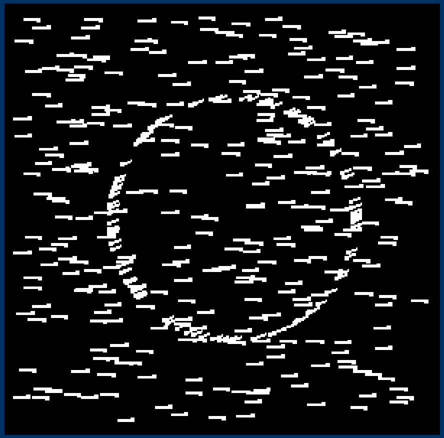


3-D view of recovered v_x velocities

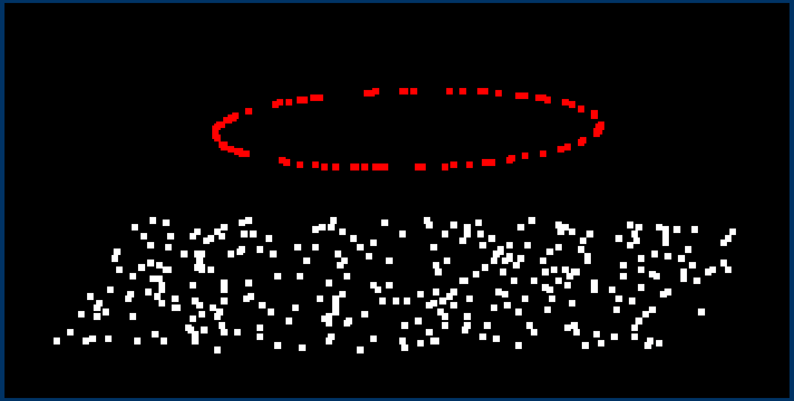
Translating Circle

Input

Handling both curves and surfaces

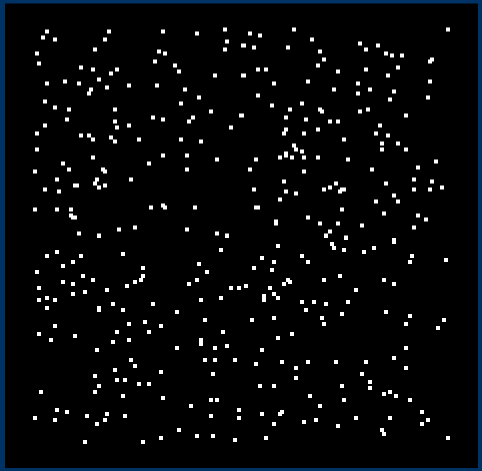


Sparse velocity field



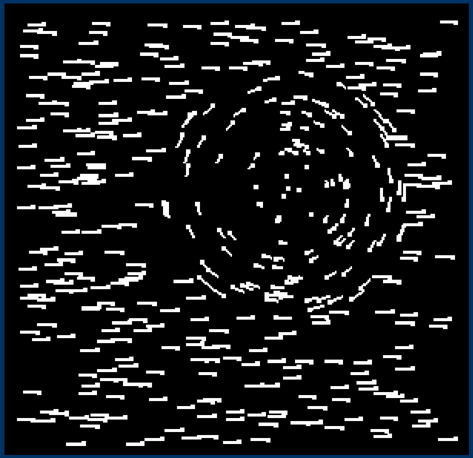
3-D view of recovered v_x velocities

Rotating Disk-Translating Background

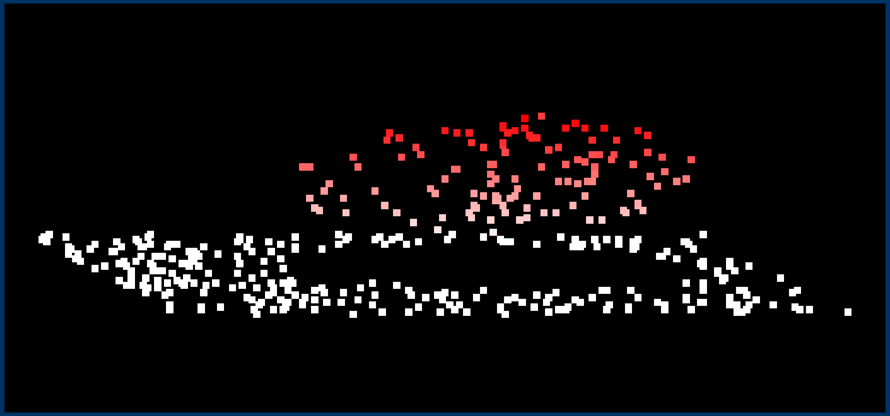


Input

No separation even in 4-D

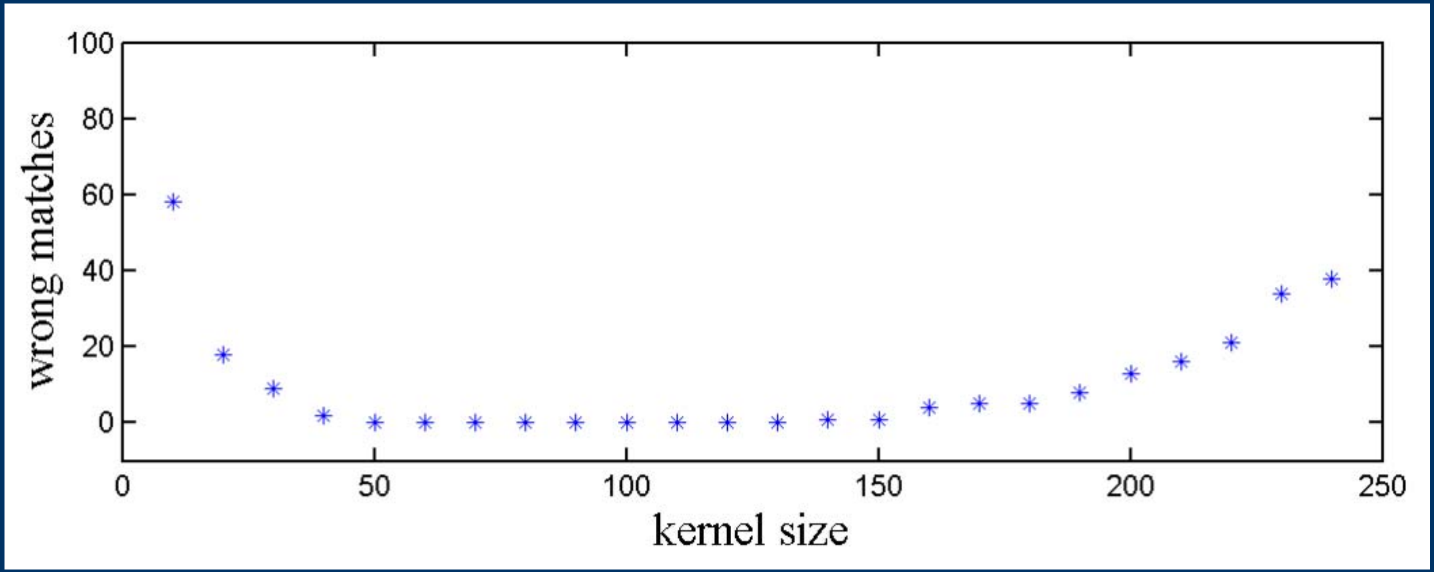


Sparse velocity field



3-D view of recovered v_x velocities

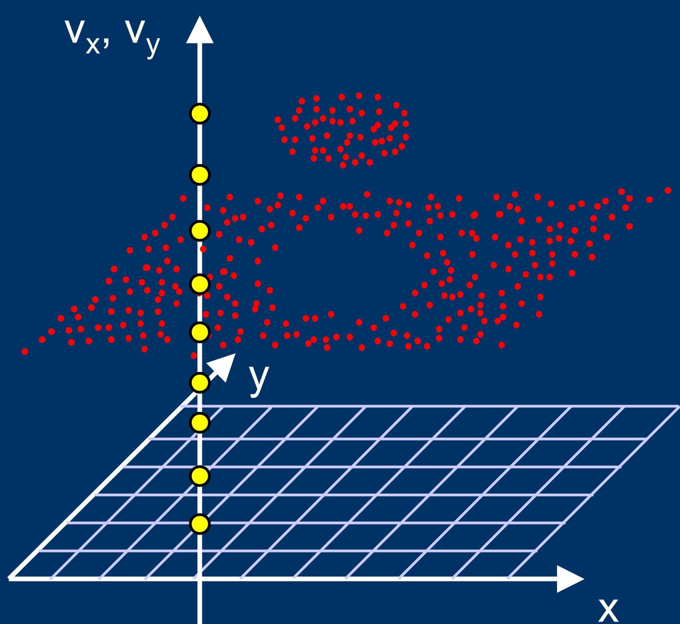
Scale Sensitivity

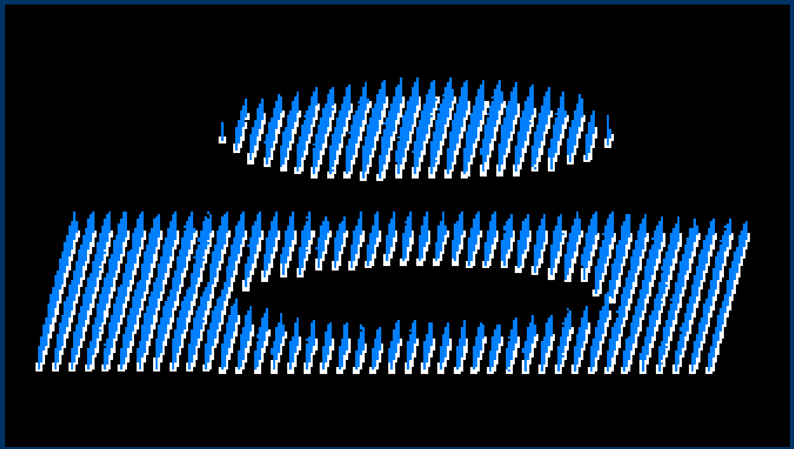


- Tested on the translating disk example
- Number of input points = 400
- Image size = 200 x 200

Densification

- At each pixel $(x\ y)$:
  - generate discrete $(v_x\ v_y)$ candidates
  - at each candidate → collect votes from the input tokens
  - use surface saliency $(\lambda_2 - \lambda_3)$ as an affinity measure
  - choose most salient candidate





Dense velocity field and layer orientations

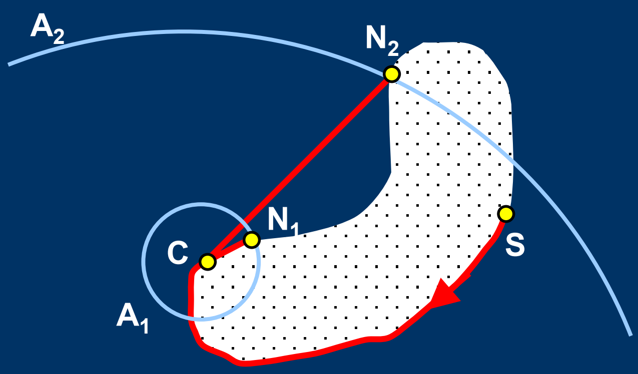
Region Grouping

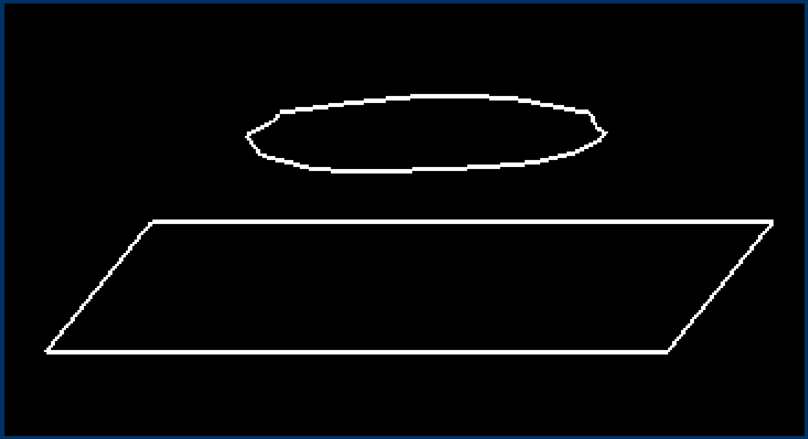
- Propagate region labels
- Criterion → smoothness of both:
  - pixel velocities ↔ distance in the $(v_x\ v_y)$ space
  - layer orientations ↔ normal vectors e_1 and e_2

Regions

Boundary Extraction

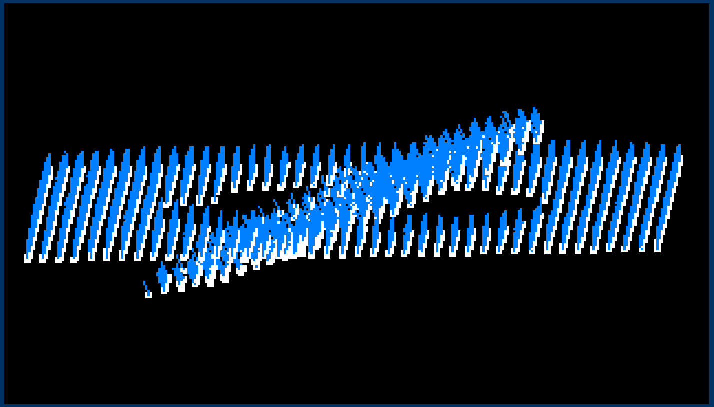
- 2-D process, that extracts a “locally convex” hull
- Irregularity – function of the scale factor
- At large scale → convex hull





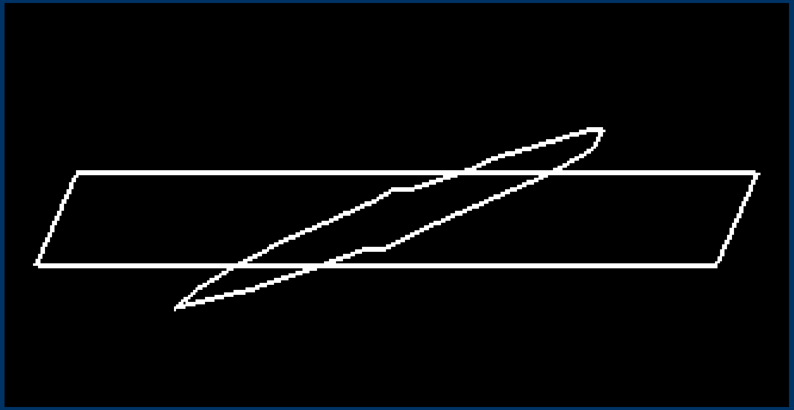
Motion boundaries

Expanding Disk



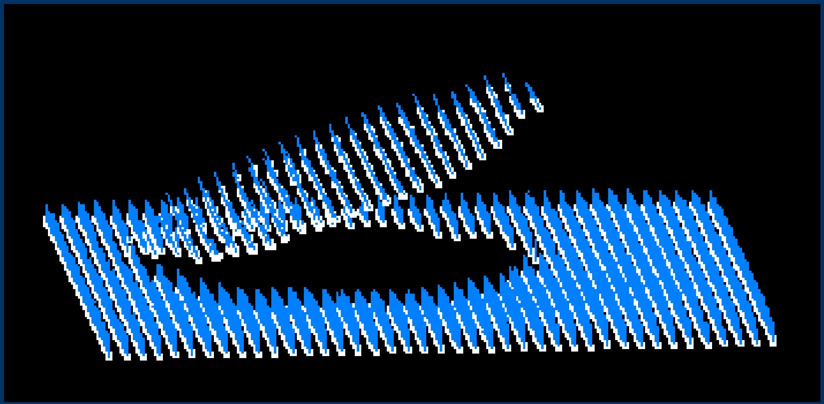
Dense velocity field

Regions

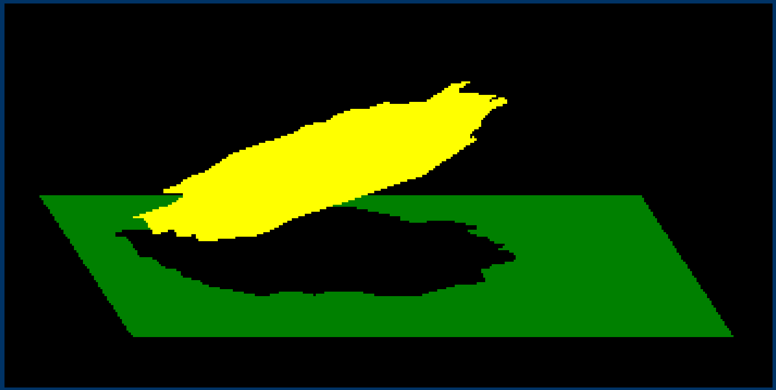


Boundaries

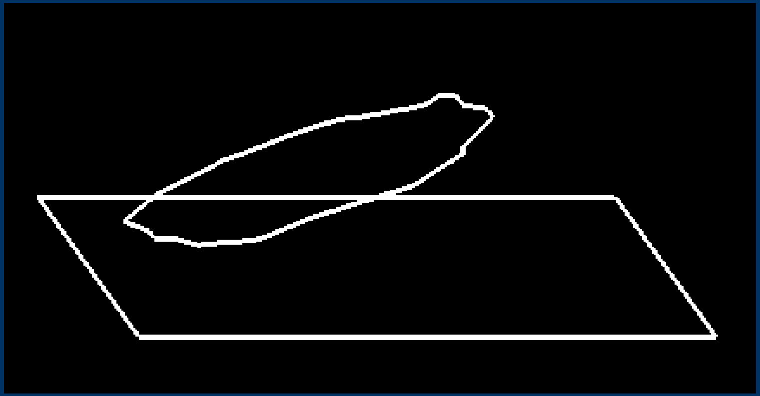
Rotating Disk-Translating Background



Dense velocity field



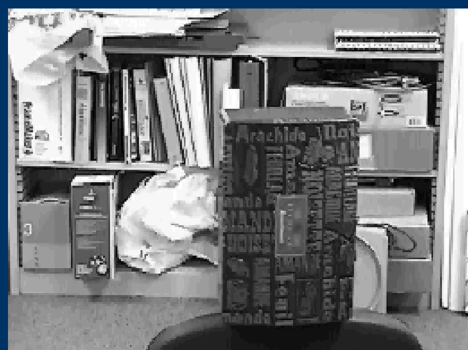
Regions



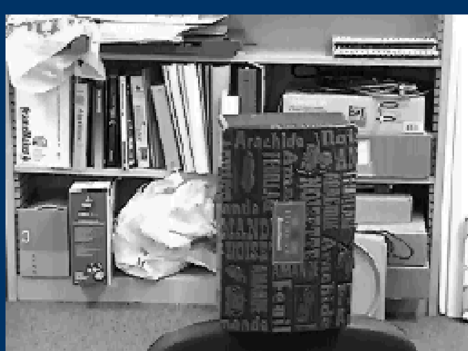
Boundaries

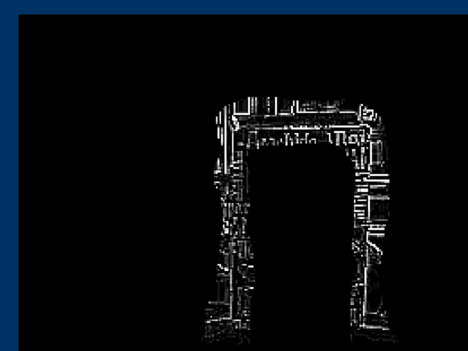
Incorporating Intensity Information

- Why not use monocular cues first?




- Augment motion with monocular cues:





Approach

- The general framework:
  - 4-D layered representation
  - Token affinity communication through voting

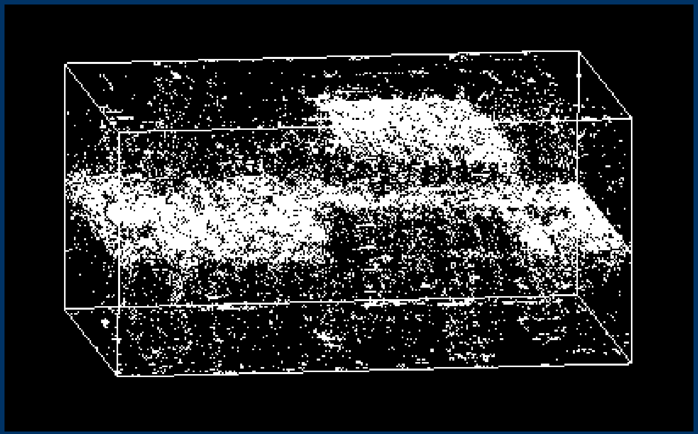
remains the same

- Issues:
  - Generation of initial candidate matches
  - Accurate boundary inference, in the presence of occlusion

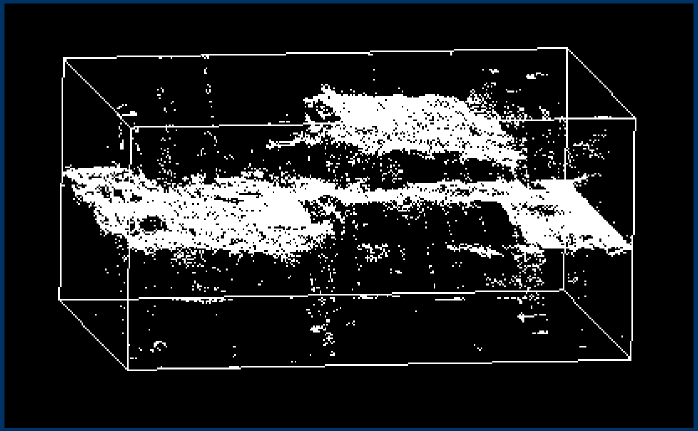
Generating Candidate Matches

- Use an intensity-based, cross-correlation procedure
- All peaks of correlation are retained as candidates
- Repeat for multiple scale values (correlation window sizes)

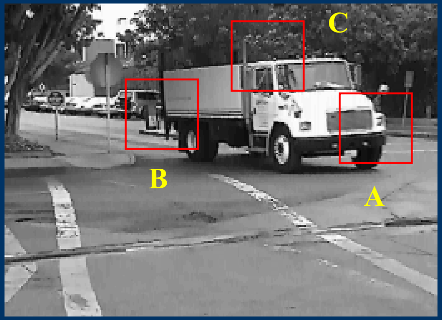
Small scale $\rightarrow$ fine detail, effective next to boundaries, noisy

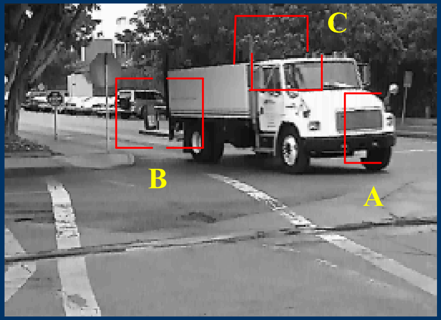


Large scale $\rightarrow$ smoother, more affected by occlusion, less noisy



Uncertainty at Motion Boundaries

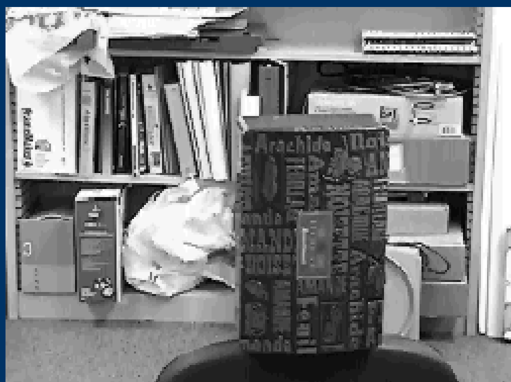




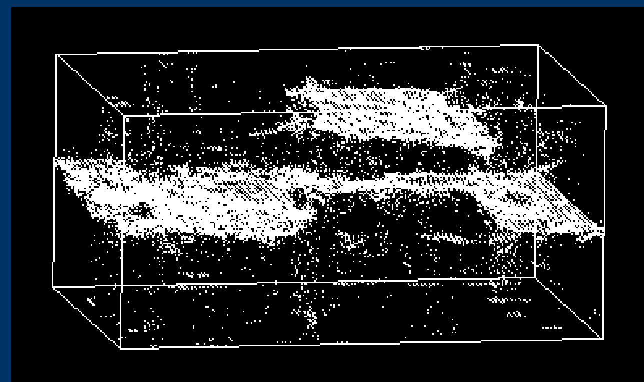
- Correlation is inherently unreliable at motion boundaries
- Non-similarity between regions
- Wrong matches may be actually consistent with the correct ones → cannot be rejected as noise

- Formulate motion analysis as a two-component process:
  - Enforce smoothness of motion, except at its discontinuities
  - Enforce smoothness of such discontinuities, aided by monocular cues

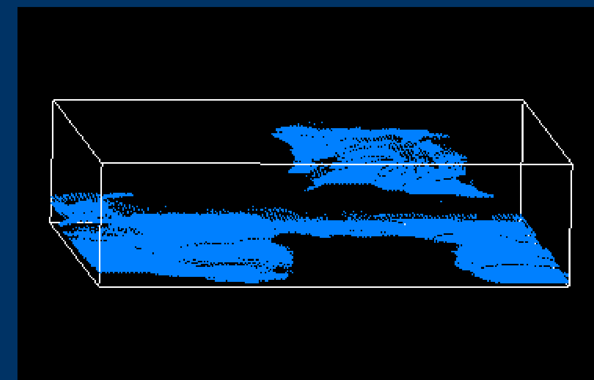
Extraction of Motion Layers



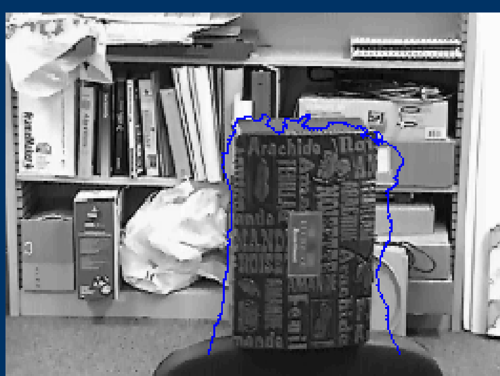
Input



Candidate matches



Dense layers



Layer boundaries

– Layers can still be over or under-extended, mainly due to occlusion

– Approach → incorporate intensity cues (edges) from original images

Boundary Saliency Map

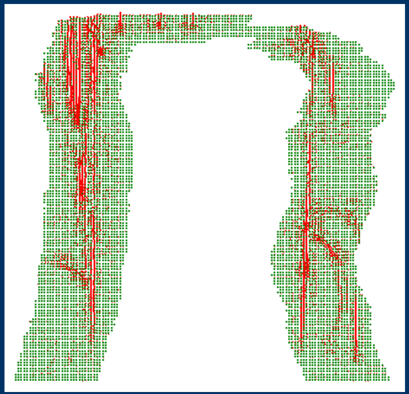
- Define a boundary saliency map in the uncertainty zones along layer boundaries
- Encode 2-D stick tensors:
  - Orientation ← gradient orientation

$$G_x(x, y) = I(x, y) - I(x-1, y)$$
$$G_y(x, y) = I(x, y) - I(x, y-1)$$

  - Saliency ← gradient magnitude

$$sal = W \cdot \sqrt{G_x^2 + G_y^2}$$

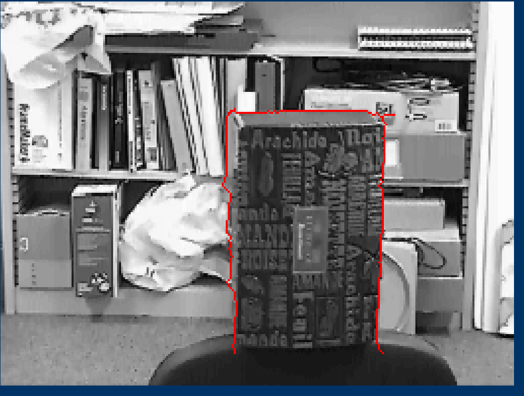
$$W = e^{-\frac{(x-x_c)^2}{\sigma_W^2}}$$



Boundary saliency map

Detecting the Boundary

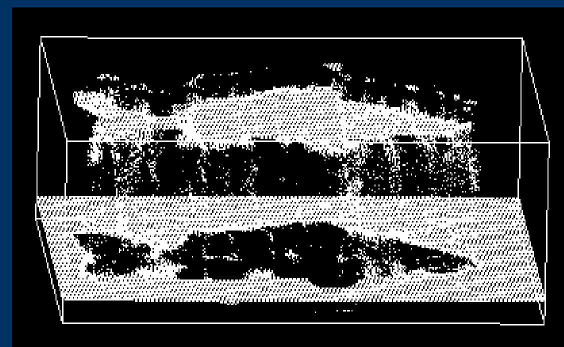
- Enforce smoothness of motion discontinuities → 2-D voting process within zones of boundary uncertainty
- After voting, grow boundary in the uncertainty zones, according to maximal curve saliency, given by $(\lambda_1 - \lambda_2)$



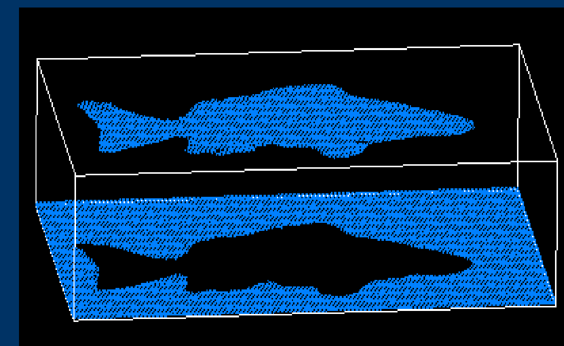
Refined boundaries

Fish Sequence (synthetic)

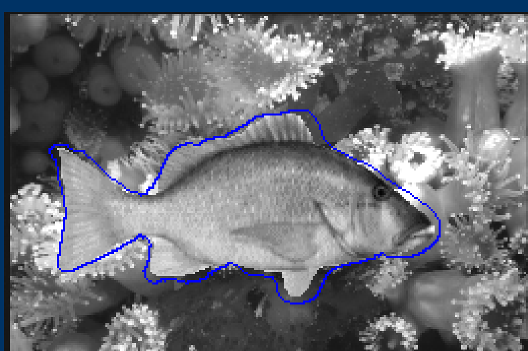
Input



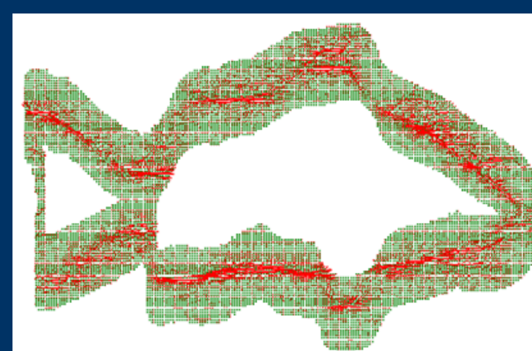
Candidate matches



Dense layers



Layer boundaries



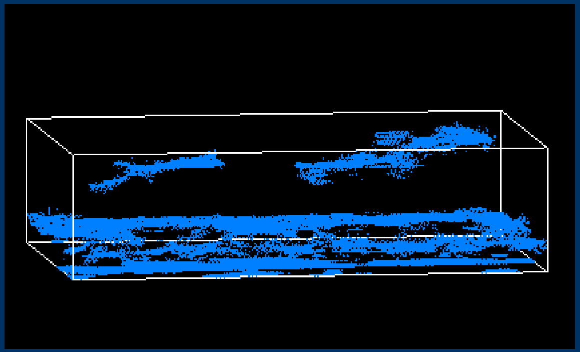
Boundary saliency map

Refined boundaries

Barrier Sequence

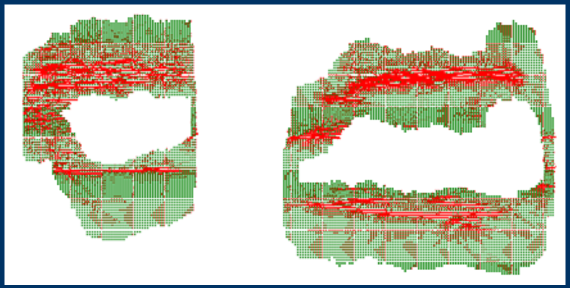
Input

Candidate matches



Dense layers

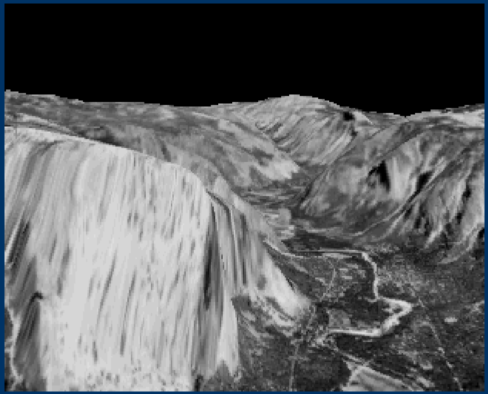
Layer boundaries



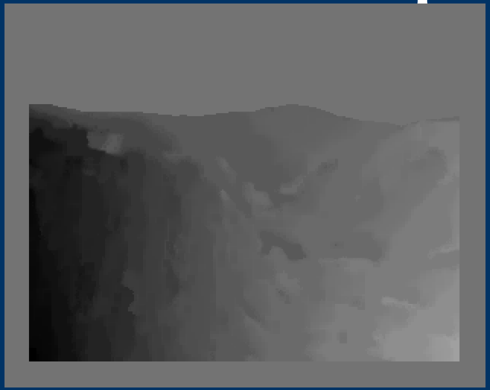
Boundary saliency map

Refined boundaries

Yosemite Sequence

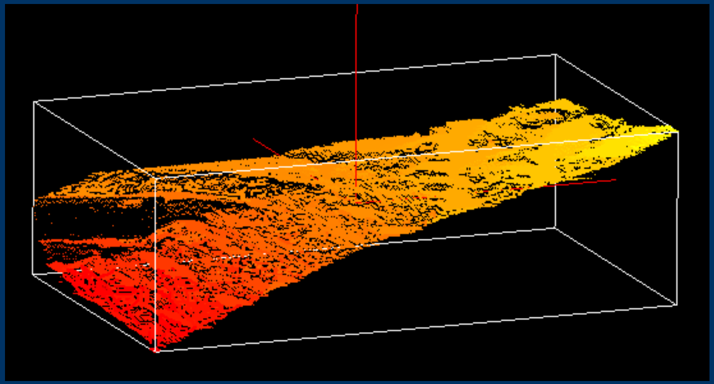


Input



x-velocities

y-velocities

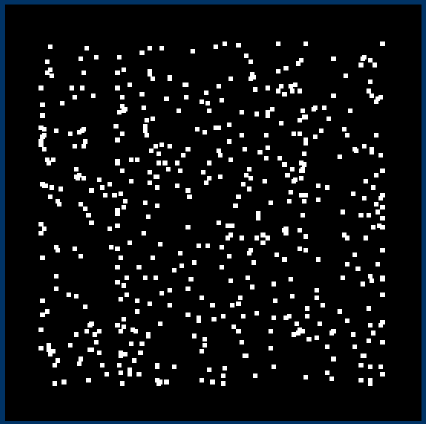


Motion layer (x-velocities)

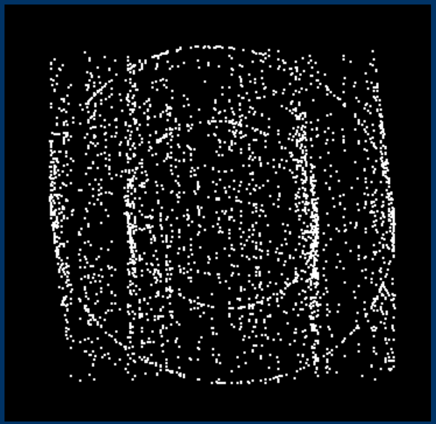
Yosemite Sequence

Technique	Average error	Standard deviation	Coverage
Nicolescu and Medioni	3.74°	4.3°	100%
Anandan	15.54°	13.46°	100%
Uras et al. (unthresholded)	16.45°	21.02°	100%
Horn and Schunck	22.58°	19.73°	100%
Lucas and Kanade ($\lambda_2 \geq 5.0$)	3.55°	7.11°	8.8%
Uras et al. ($det(H) \geq 2.0$)	3.75°	3.44°	6.1%
Fleet and Jepson ($\tau = 2.5$)	4.29°	11.24°	34.1%
Fleet and Jepson ($\tau = 1.25$)	4.95°	12.39°	30.6%
Lucas and Kanade ($\lambda_2 \geq 1.0$)	5.20°	9.45°	35.1%
Uras et al. ($det(H) \geq 1.0$)	5.97°	11.74°	23.4%
Heeger	11.74°	19.0°	44.8%

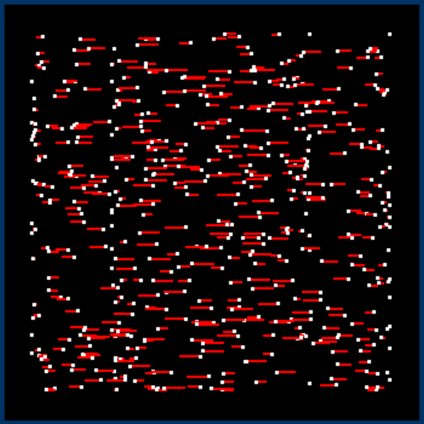
Cylinders Sequence



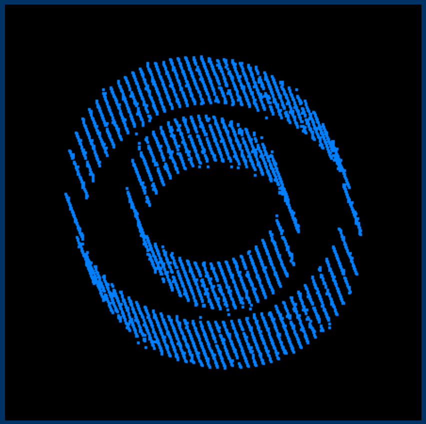
Input



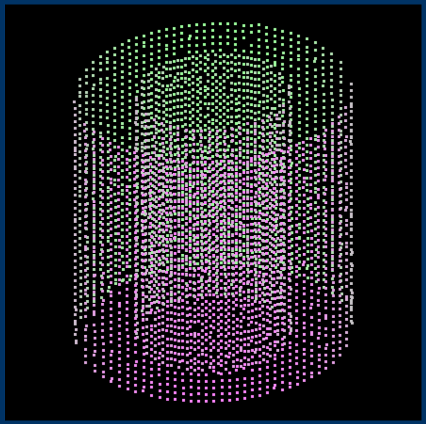
Candidate matches



Velocities



Dense layers



3-D structure