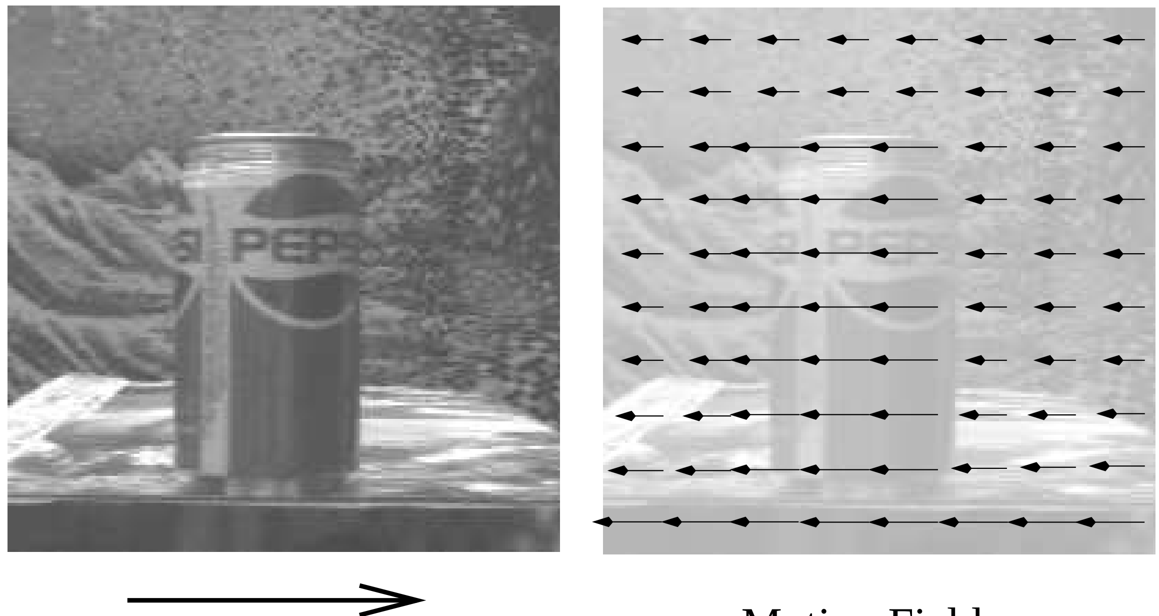
The Motion Field and Optical Flow

Motion Field: The image motion of fixed scene points.

Optical Flow: Approximation of the motion field obtained by tracking "stuff" (image texture) across two or more frames.



Camera Motion

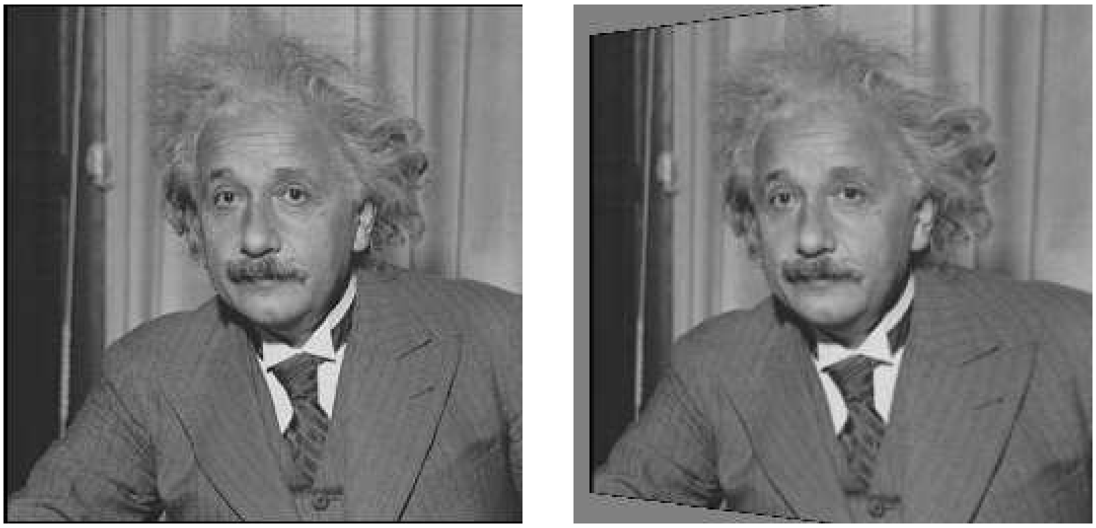
Motion Field

Difficulties in optical flow estimation:

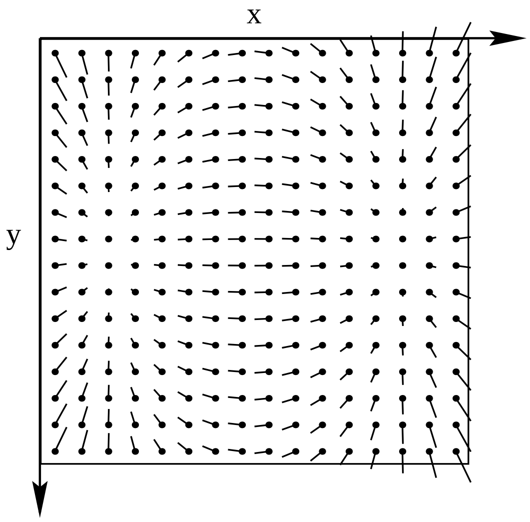
- Variation in shading and reflections.
- Occlusion and transparency dictate a discontinuous or multiple valued flow field.
- Insufficient image texture for local 2D tracking.
- Deformation of image texture.
- Large displacements are common between successive frames.

Rubber Sheet Deformation

Consider the image pair:



The deformation is a 2D displacement field, $\vec{v}(\vec{x})$:



Here:

- $\vec{W}(\vec{x}) = \vec{x} + \vec{v}(\vec{x})$ is the image warp.
- $I_r(\vec{W}(\vec{x})) = I_l(\vec{x})$ describes the "rubber sheet" image deformation.

Representation of Smooth Deformation Fields

Pure Translation:

$$\vec{W}(\vec{x};\vec{a}) = \vec{x} + \vec{a}, \quad \text{with } \vec{a} \in \Re^2.$$

Affine Warp:

$$\vec{W}(\vec{x},\vec{a}) = \vec{a}_0 + A\vec{x}, \quad \text{with } \vec{a} \in \Re^6.$$

Provides a good approximation for the motion of objects with

- Small depth variation.
- Small angular extents within the image.

Rational Warp:

$$\vec{W}(\vec{x};\vec{a}) = \{\vec{a}_0 + A\vec{x}\} \,/\, \left[1 + \vec{b}\cdot\vec{x}\right] \quad \text{with } \vec{a} \in \Re^8.$$

Provides a good approximation for

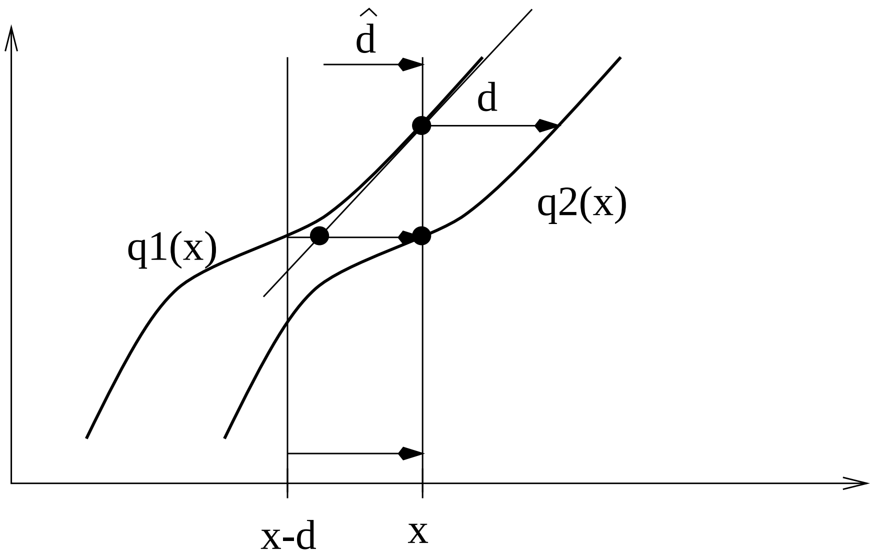
- Thin structures (nearly planar surfaces).
- No restriction on angular extent.

Can also consider patchwise representations of smooth flow, or splines, RBFs, and so on.

Estimating 1D Signal Displacement

Given two signals, $q_1(x)$ and $q_2(x)$, the displacement d satisfies

$$q_2(x) = q_1(x - d).$$



Given information about q_1 and q_2 at only one point x, we can use a **local linear model:**

$$q_1(x + r) = q_1(x) + r\frac{dq_1}{dx}(x) + O(r^2).$$

Neglecting $O(r^2)$ terms, d can then be approximated by

$$\hat{d} = [q_1(x) - q_2(x)]/\frac{dq_1}{dx}(x).$$

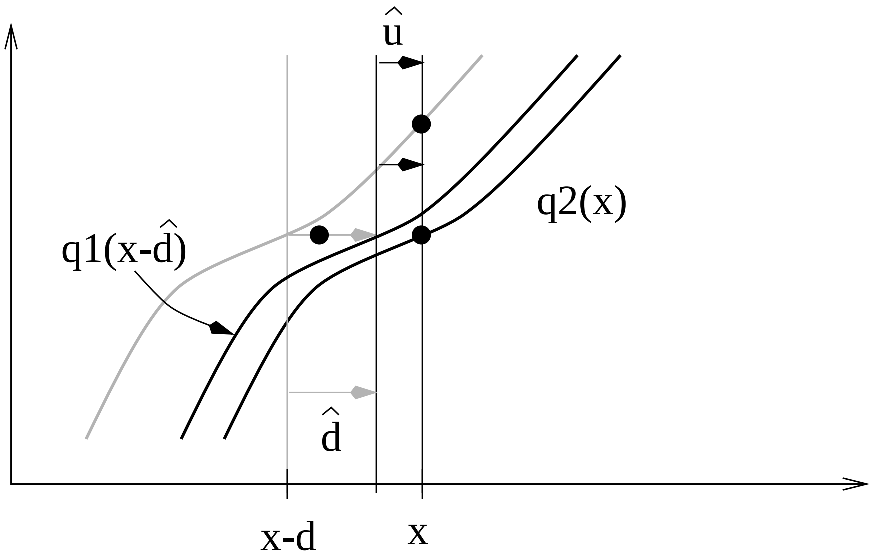
To leading order, the error is

$$|\hat{d} - d| \leq \frac{1}{2}\hat{d}^2|\frac{d^2q_1}{dx^2}(x)|/|\frac{dq_1}{dx}(x)|.$$

Iterative Displacement Matching

Given an estimate $\hat{d}$ for the displacement, consider estimating the update, u:

$$q_2(x) = q_1(x - \hat{d} - u).$$



The new estimate for the displacement is $\hat{d} + \hat{u}$ where

$$\hat{u} = [q_1(x - \hat{d}) - q_2(x)] / \frac{dq_1}{dx}(x - \hat{d}).$$

The error in this new estimate is $O(u^2)$.

Further iteration gives Newton's method applied to

$$g_1(x - d) - g_2(x) = 0.$$

The approximation error $O(u^2)$ vanishes upon convergence (the displacement error will then be dominated by errors in $q_1(x)$ and $q_2(x)$).

Choice of $q(x)$

To get an accurate initial displacement estimate, $\hat{d}$, over a wide range of image displacements, we require the image property $q(x)$ to be essentially linear over that range.

Image Intensities (Grey-levels): Linearity assumption is satisfied for smoothly shaded regions but, in texture areas, method may be expected to work only for displacements of a pixel or so.

- Use a low-pass image pyramid.
- Coarse scales provide rough estimate of displacement for shifts up to a pixel or so *in the coarse sampling grid.*
- The coarse estimate is then used to initialize an estimate on the next finer scale.
- The finer scale can be expected to correct this value by up to a pixel or so in it's sampling grid.
- And so on...

This is called **coarse-to-fine image matching**.

Wavelet Filter Responses (Phase): Their bandpass characteristics allow us to use linear models of phase over larger distances than a (subsampled) pixel.

Estimation of Smooth Deformation Fields

Brightness Constancy Assumption: For an image region $\mathcal{R}$, seek a warp $\vec{W}(\vec{x};\vec{a})$ that minimizes the rubber sheet matching error

$$E(\vec{a}) = \sum_{\vec{x}\in\mathcal{R}} \rho(I(\vec{W}(\vec{x};\vec{a}),t) - I(\vec{x},t-1)).$$

Here

- $I(\vec{x},t)$ and $I(\vec{x},t-1)$ are successive image frames.
- $\rho(e)$ is the *estimator* (eg. a simple choice is $\rho(e) = e^2$).
- Minimizing $E(\vec{a})$ is a nonlinear optimization problem for $\vec{a}$.

Differential Formulation

For near-identity warps, that is $\vec{W}(\vec{x};\vec{a}) \approx \vec{x}$, a computationally convenient approximation to $E(\vec{a})$ is

$$\tilde{E}(\vec{a}) = \sum_{\vec{x}\in\mathcal{R}} \rho(\nabla I(\vec{x},t)\cdot\vec{v}(\vec{x};\vec{a}) + \frac{\partial}{\partial t} I(\vec{x},t)).$$

Here

- $\vec{v}(\vec{x};\vec{a}) = \vec{W}(\vec{x};\vec{a}) - \vec{x}$.
- $\nabla I = (\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y})$ is the spatial gradient of the image.

Each individual differential constraint,

$$\nabla I(\vec{x},t)\cdot\vec{v}(\vec{x};\vec{a}) + \frac{\partial}{\partial t} I(\vec{x},t) = 0$$

is called a "motion constraint equation."

Iterative Rewarping: For accurate displacement estimates, the computed minimum, say $\vec{a}_\mu$, is used to warp the frame at time t

$$I'(\vec{x},t) = I(\vec{W}(\vec{x};\vec{a}_\mu),t).$$

Matching is performed again on the warped pair $(I'(\vec{x},t), I(\vec{x},t-1))$. This also compensates for changes of scale, orientation, and so on.

Motion Constraint Line

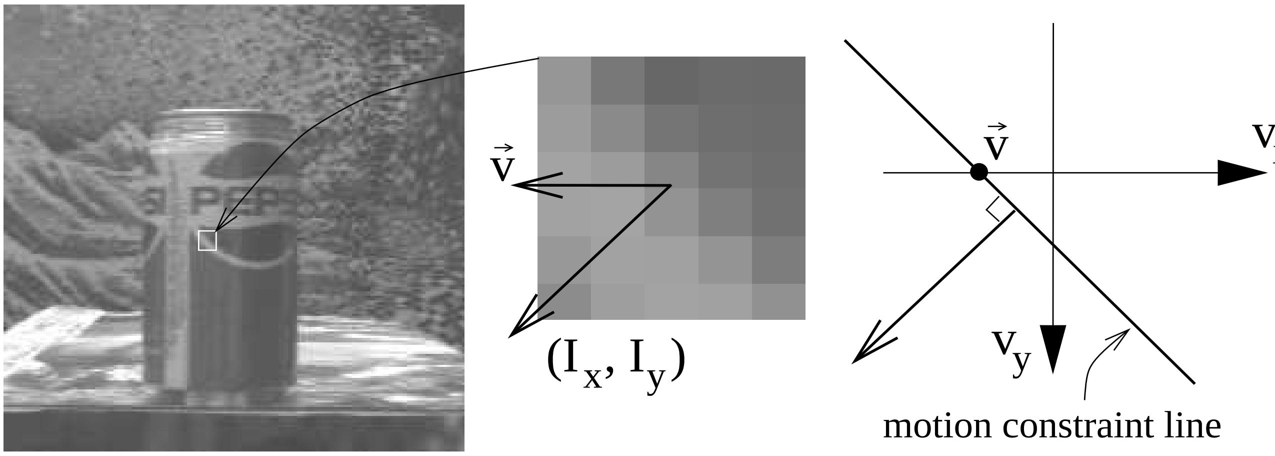
The motion constraint equation has the form:

$$\vec{c} \cdot (v_x, v_y, 1) = 0$$

Here

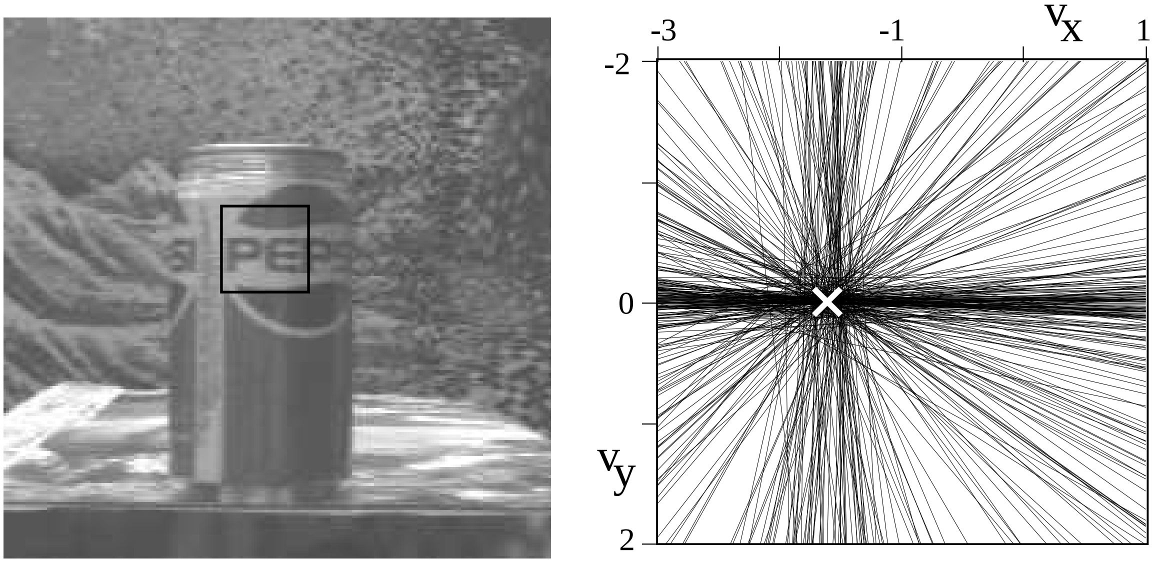
- $\vec{c} = (I_x(\vec{x}, t), I_y(\vec{x}, t), I_t(\vec{x}, t))^T$ can be obtained from image measurements.
- $\vec{v}(\vec{x}, t) = (v_x, v_y)^T$ is the unknown 2D image velocity at $(\vec{x}, t)$.

Geometrically, one such constraint determines a line in $\vec{v}$-space, perpendicular to $(I_x(\vec{x}, t), I_y(\vec{x}, t))$.

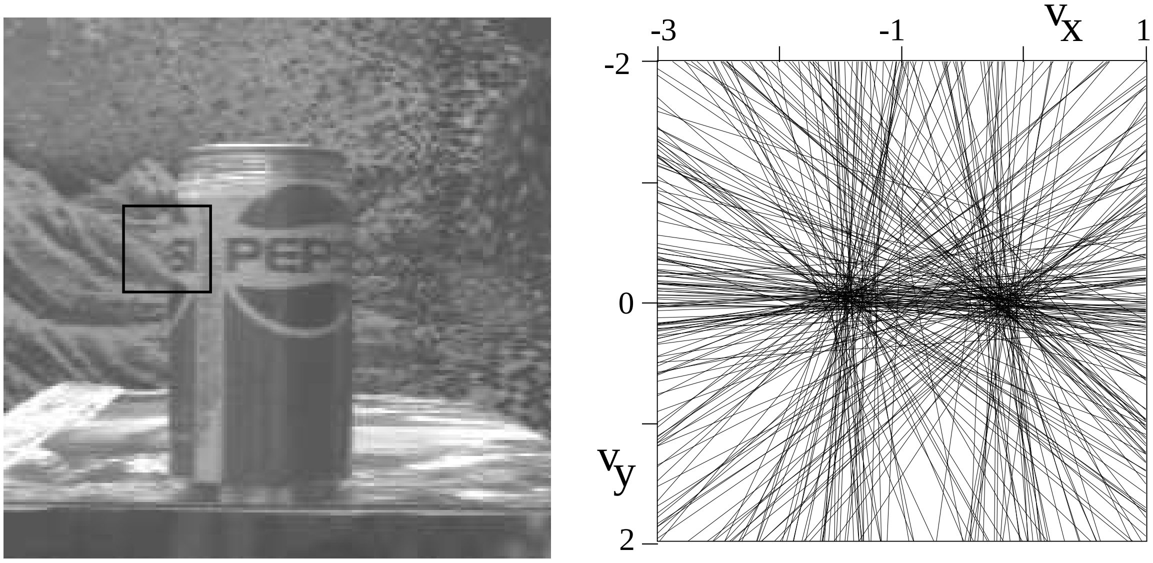


Motion Test Cases

1. Constant velocity within an image patch:



2. Occlusion boundary within an image patch:

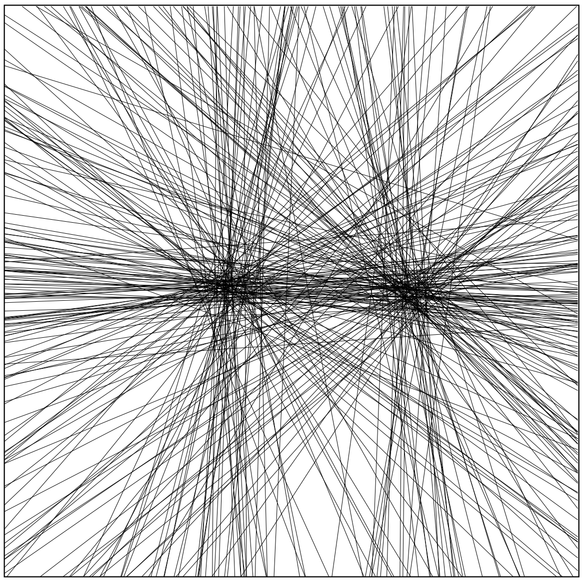


In case 2, least squares estimation (i.e. $\rho(e) = e^2$) is not appropriate.

The Aperture Problem

Computation of image motion requires the integration of several constraints from within an aperture.

Example: Patch contains an occlusion boundary:



Intersection of constraint lines

$$\vec{c}_k \cdot (v_x, v_y, 1) = 0$$

gives motion estimate(s).

Aperture Problem (Cont.)

The aperture must be chosen:

- sufficiently $\boxed{\text{LARGE}}$ to collect enough constraints,

and

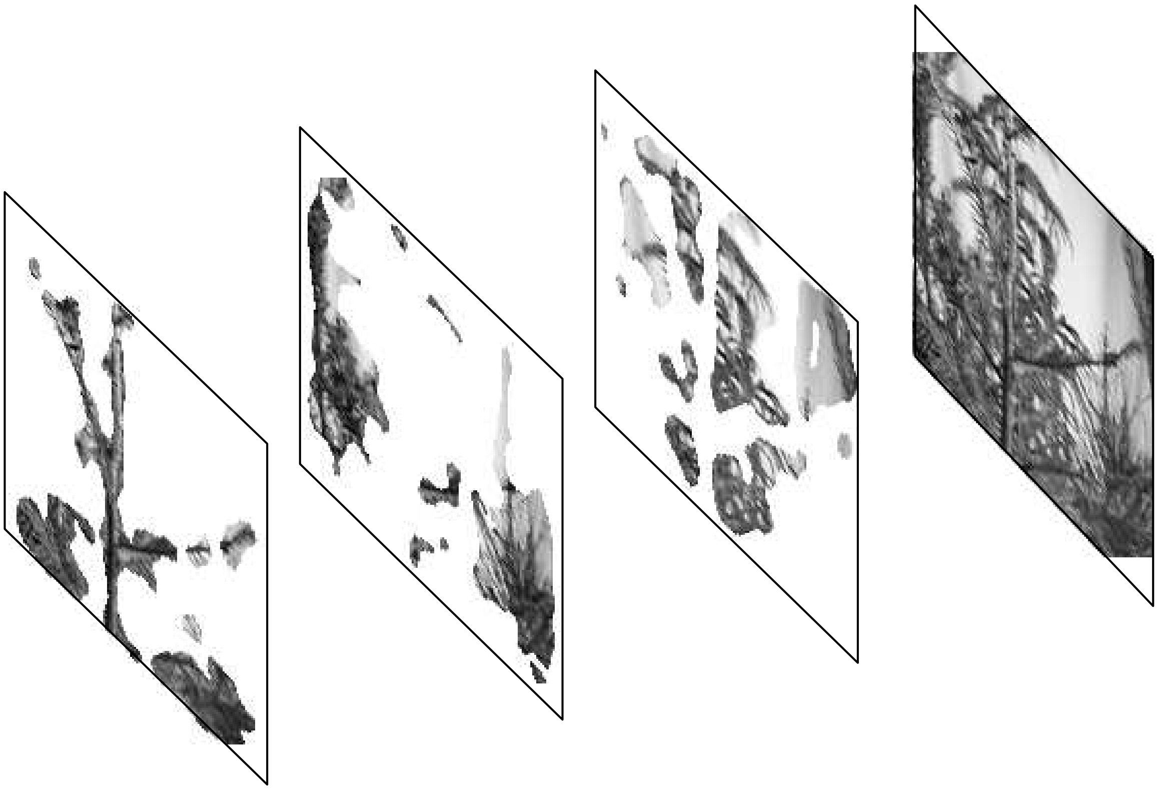
- sufficiently $\boxed{\text{small}}$ (with a suitable shape) to collect constraints from (predominantly) a single motion.

Sources of multiple motions:

- Occlusion boundaries
- Reflections
- Transparency
- Shadows

Fragmented Occlusion

Cannot expect to tune the aperture shape to complex fragmented regions.



Foreground *Middle* *Back* *Image*

Integration of Motion Constraints

Method for integrating motion constraints must accommodate multiple motions within a single aperture.

Possible approaches:

- Regularization with motion boundaries.

  (Black & Anandan – CVPR'91)

- Stochastic multi-layer segmentation

  (Darrell & Pentland – Motion Workshop '91)

- Robust regression and regularization.

  (Black & Anandan – ICCV'93)

- Mixture models

  (Jepson & Black – CVPR'93)

- Skin and bones

  (Ju, Black & Jepson – CVPR'97)

Mixture Models for Optical Flow

Basic idea: Motion constraints come from a mixture of simple probability distributions:

- One simple distribution for each of several different *layers*.
- Each layer corresponds to a smooth motion field $\vec{W}(\vec{x}; \vec{a}_l)$.
- A background distribution for *outliers* is included.
- Each motion constraint is assigned an *ownership probability* for belonging to each layer and to the outlier distribution.

The *Expectation-Maximization (EM) algorithm* is used to obtain a maximum likelihood fit for the parameters of the mixture model, including the warp parameters $\vec{a}_l$ for each layer.

The introduction of an outlier layer is roughly equivalent to using a *robust estimator* for $\rho(e)$ (instead of the squared error e^2).

Specification of a Mixture Model

Probability of observing motion constraint $\vec{c}_k$

$$p(\vec{c}_k | \vec{x}_k, \vec{m}, \vec{a}_1, \ldots, \vec{a}_N) = \sum_{n=0}^{N} m_n p_n(\vec{c}_k | \vec{x}_k, \vec{a}_n),$$

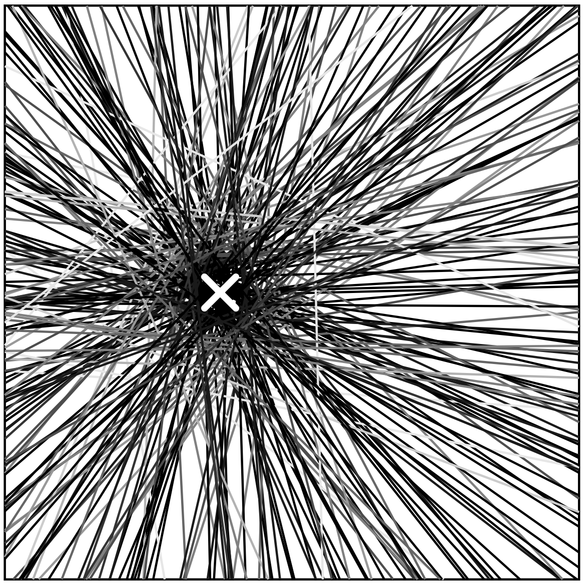
where

- $p_n(\vec{c}_k | \vec{x}_k, \vec{a}_n)$ = probability of observing motion constraint $\vec{c}_k$ given it comes from the n^{th} layer at spatial location $\vec{x}_k$.
- m_n is the mixture proportion for the n^{th} layer
- $\vec{a}_n$ = parameters of n^{th} layer

Mixtures of Gaussian Distributions

Gaussian distribution of component velocities for n^{th} layer:

$$p_n(\vec{c}|\vec{x}, \vec{a}) = \text{Gaussian}(\vec{c} \cdot \vec{v}(\vec{x}, \vec{a}) \mid 0, \sigma_v)$$



Models for n^{th} layer:

- Constant velocity within patch:

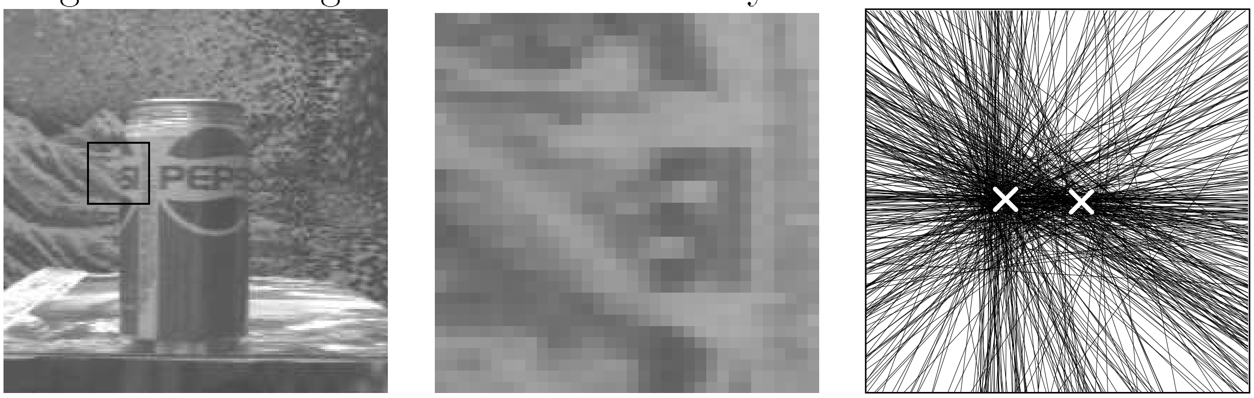
  $\vec{a}$ two dimensional: $\vec{v}(\vec{x}, \vec{a}) = \vec{a}$

- Affine motion within patch:

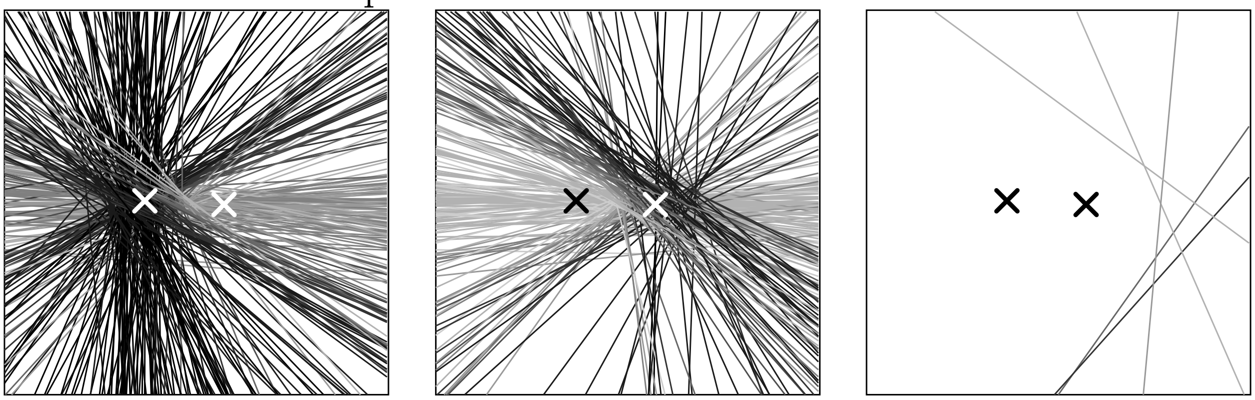
  $\vec{a}$ six dimensional

Representation for Occlusion Example

A region containing an occlusion boundary:

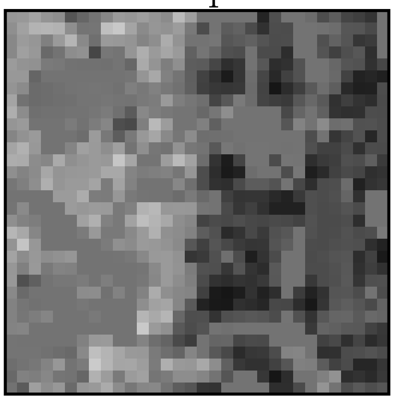


Constraint Ownership:



Layer #1 | Layer #2 | Outliers

Backprojected Constraint Ownership:



Fitting Mixture Models to Data

Maximum likelihood fits using the EM-algorithm

Iterate:

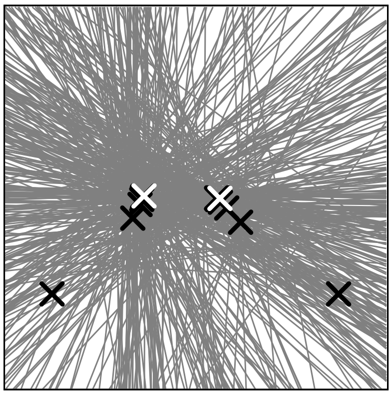
E-step:

- For each constraint $\vec{c}_k$, estimate the "ownership probability" for each of the layers.
- Simple computation involving the evaluation of $p_n(\vec{c}_k)$

M-step:

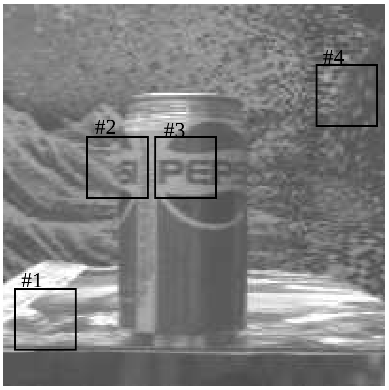
- Use the ownership probabilities as weights in a linear least squares problem for the layer parameters $\vec{a}_n$.

Convergence Behaviour:

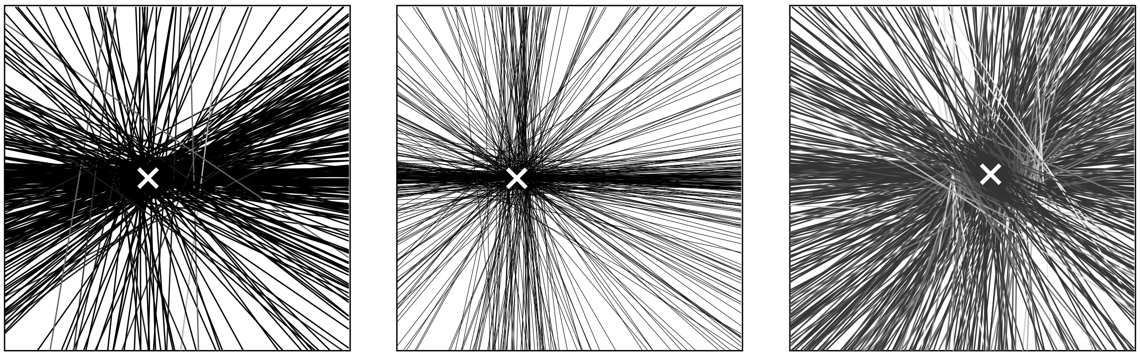


Results for Additional Test Patches

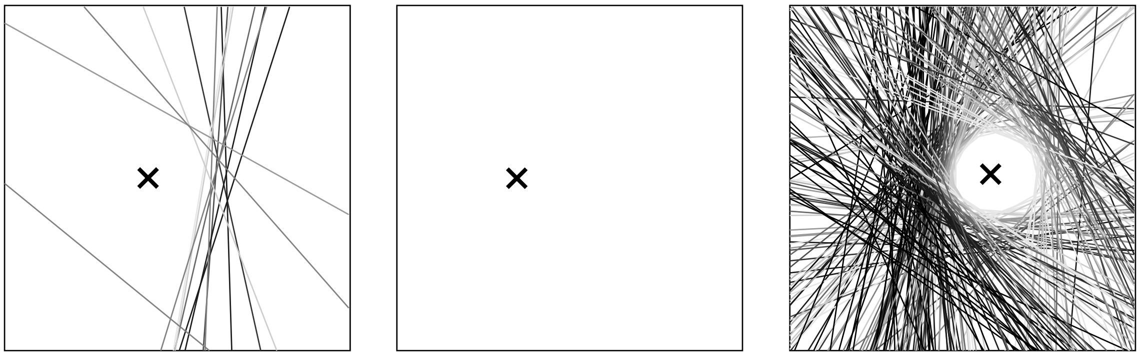
Test Patches:



Motion constraints for regions #1, #3 and #4:



Outliers for regions #1, #3 and #4:



Mixtures of Affine Models

Affine Model: Optical flow $\vec{v}(\vec{x})$ given by a *linear* function of image coordinates $\vec{x}$:

$$\vec{v}(\vec{x}, \vec{a}) = \vec{a}_0 + A\vec{x}$$

Shallow Structures

- Small depth variation relative to mean distance.
- Small angular extent of view.

Optical flow for a shallow structure is well approximated by an affine model.

M-Step requires solution of 6D linear system for $\vec{a}$, for each component of the mixture distribution.

Mixtures of Rational Models

Rational Model: Image position in new frame, $\vec{X}(\vec{x})$, given by a *rational* function of image coordinates $\vec{x}$:

$$\vec{X}(\vec{x}; \vec{a}, \vec{b}) \quad = \quad \left\{ \frac{\vec{a}_0 + A\vec{x}}{1 + \vec{b} \cdot \vec{x}} \right\}$$

Thin Structures

- Nearly planar 3D structure.
- Deviation from a plane is small relative to distance at that point.
- No restriction on angular extent of view.

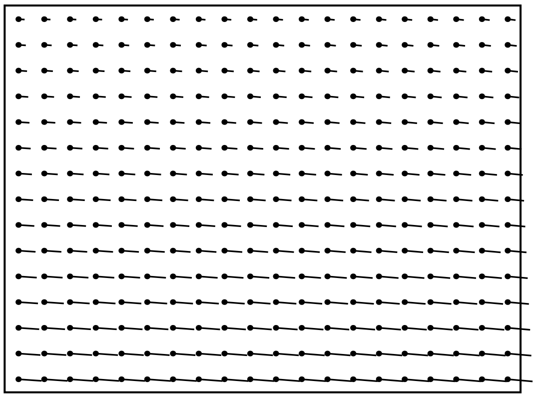
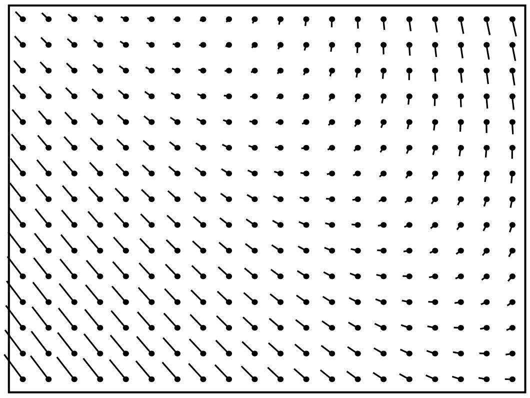
Image displacement mapping for a thin structure is well approximated by a rational model (exact for a planar surface).

M-Step requires solution of 8D nonlinear system for $\vec{a}$ and $\vec{b}$, for each component of the mixture distribution.

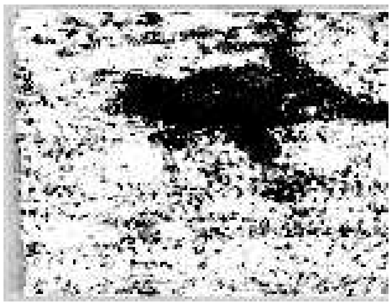
Layered Motion Model

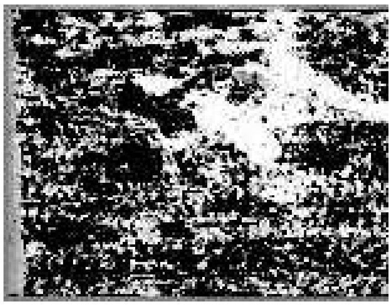
Two successive image frames:


Two affine layers fit to the dominant motion components:

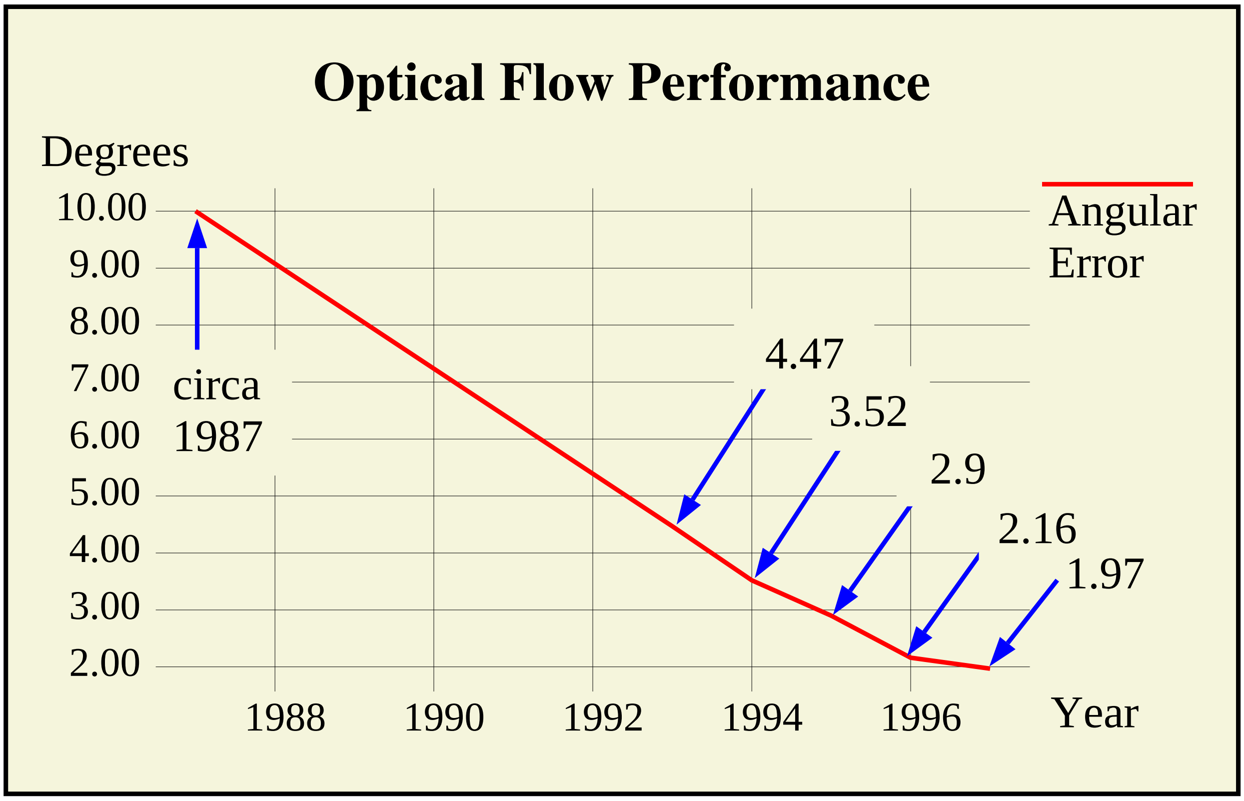
Ownership maps for the two motion layers plus outliers:

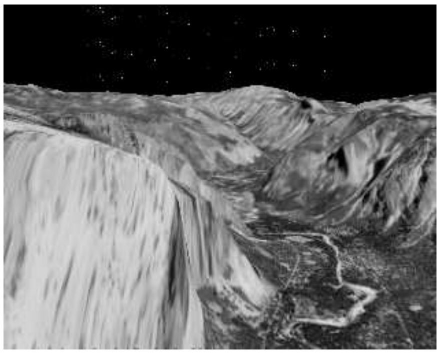


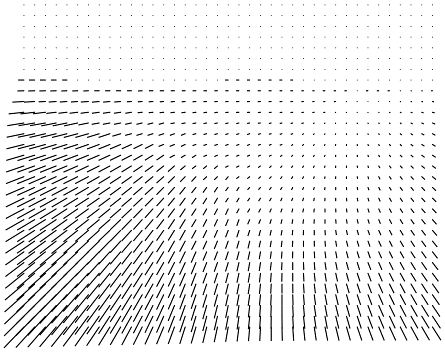


Layer 1 | Layer 2 | Outliers

Yosemite Sequence







Ju, Black & Jepson, 1997

Summary

Mixture models provide a robust and reliable tool for estimating motion or displacement parameters.

The approach:

- Relaxes the single motion assumption.
- Recovers multiple affine or rational motions.
- Detects outliers.

Current research issues:

- Determining the number and type of the layers.
- Dealing with leverage problems.
- Filling in motion for image patches with 1D structure.
- The use of spatial priors for layer ownerships.
- Coupling motion with texture, colour, or stereo information.