Robust Visual Motion Analysis: Piecewise-Smooth Optical Flow and Motion-Based Detection and Tracking

Ming Ye
Electrical Engineering, University of Washington
mingye@u.washington.edu

What Is Visual Motion

- 2D image velocity
  - 3D motion projection
  - Temporal correspondence
  - Image deformation

- Optical flow
  - An image of 2D velocity
  - Each pixel \( V_s(x,y) = (u_s,v_s) \)
  - \((x,y,t) \mapsto (x+u,y+v,t+1)\)

Structure From Motion

Rigid scene + camera translation
Estimated horizontal motion
Depth map

Scene Dynamics Understanding

- What’re moving? How?
  - Surveillance
  - Event analysis
  - Video compression

Motion smoothness
Target Detection and Tracking

A tiny airplane --- only observable by its distinct motion

Tracking results

Image Distortion Measurement

- Image deformation
  - Measure it. Remove it.
  - Image-based rendering

Research Areas

- Structure from motion
- Scene dynamics analysis
- Object detection and tracking
- Video compression
- Image/video enhancement
- Image-based rendering
- Visual motion estimation

Outline

- Optical flow estimation
  - Background
  - A local method with error analysis
  - A Bayesian approach with global optimization
- Motion-based detection and tracking
Optical Flow Estimation

Basics

- **Template matching**
- **Assumptions:**
  - Brightness conservation
  - Flow smoothness
- **Difficulties:**
  - Aperture problem (local information insufficient)
  - Outliers (motion boundaries, abrupt image noise)

Previous Work (1/2)

- **Brightness conservation**
  - Matching-based \( I(x, y, t) = I(x + u, y + v, t + 1) \)
  - Gradient-based \( I_x + I_y + I_t = 0 \) (OFC)

- **Flow smoothness**
  - Local parametric \( \frac{d}{dx} (x, v) = (u, 0) \) \([\text{Lucas-Kanade 81}][\text{Haralick-Lee 83}]\)
  - Global optimization \( \min_{u, v} \sum_{(x, y)} \left( I_x - u \right)^2 + \left( I_y - v \right)^2 \) \([\text{Horn-Schunck 81}]\)

Previous Work (2/2)

- **Handle motion discontinuities & Outliers**
  - Robust statistics \( \arg \min \sum \left( |I(x, y, t)| + \sigma_y \right) + \lambda \sum \left( |(v, u)| + \sigma_v \right) \) \([\text{Black-Anandan 96}]\)
  - Many others
- **Higher-level methods**
- **Problems:**
  - Gradient calculation
  - Global formulation; \( \sigma_x, \sigma_y, \lambda \) values?
  - Computational complexity
Two-Stage Robust Optical Flow Estimation with Error Propagation

A Local Approach

Method

- **2-stage regression (LS)** [Haralick-Lee 83, Ye-Haralick 98]
- **Previous: robust OFC only**
- **2-stage-robust adaptive scheme** [Ye-Haralick 00]

Error Analysis

- **Covariance propagation** [Haralick 96]
  - (Approx.) linear system + small errors
- **Previous work**
  - | Image noise var. | EIV | OFC corr. |
  - | Simoncelli 91 | Yes | Yes | No |
  - | Szeliski 89 | Yes | No | No |
  - | Nagel 94 | No | Yes | Yes |
  - | Ye-Haralick 98 | Yes | Yes | Yes |

- **New: reject outliers first**
Results

- A simple motion boundary detector
  - EIV + OFC correlation
  - Simoncelli equiv
  - LS
  - Our old
  - Robust
  - Our new

- Error analysis: why bother
  - Accurate uncertainty is just as important
  - Uncertainty is anisotropic, varies from site to site

Problem Statement

Assuming only brightness conservation and piecewise-smooth motion, find the optical flow to best describe the intensity change in three frames.

MAP/MRF Formulation

- Maximum A Posterior Criterion:
  \[
  \hat{V} = \arg \max_P (V / D) = \arg \max_P (P(D | V) P(V))
  \]

- Prior: Markov Random Fields
  - Neighborhood system: 8-connected \( N^8 \), pairwise
  - Gibbs distribution equivalent
  \[
  P(V) = \exp(-E_s(V)) / Z, \quad E_s(V) = \sum_{x \in N^8} \rho(V_x - V_{x'}, \sigma_x)
  \]

- Likelihood: exponential
- Global optimization problem

Estimating Piecewise-Smooth Optical Flow with Global Matching and Graduated Optimization

A Bayesian Approach
Global Energy Design

- **Global energy** \( E = \sum_{s \in \text{sites}} E_g(V_s) + E_v(V_s) \)
- **Matching error** \( E_g(V_s) = \rho(e_g(V_s), \sigma_s) \)
  - **Warping error** \( e_g(V_s) = \min(|I(V_s) - I_s|, |I'(V_s) - I_s|) \)
- **Smoothness error** \( E_v(V_s) = \frac{1}{8} \sum_{s \in \text{sites}} \rho(|V_s - V_{s'}|, \lambda_s) \)

3-Frame Matching Without aliasing, all pixels in a frame are visible in the previous or the next frame.

Error Function \( \rho(x, \sigma) : \)

- A distribution with fatter tails
- An error norm less drastic than L2
  - Robust against outliers
  - Simultaneous segmentation
    - Smoothness outliers = motion discontinuities
- Use Geman-McClure for redescending & normalization
  \[ \rho_1(x, \sigma) = \frac{1}{\sigma^2 + x^2}, \quad \rho_2(x, \sigma) = \rho'(x, \sigma) = \frac{2\sigma}{(\sigma^2 + x^2)} \]

Advantages

- Compare with [Black-Anandan 96]
  \[ \arg \min_{\{u_b, v_b, I_b, \sigma_s\}} \sum_{s \in \text{sites}} \{ \rho(u_b - u_s, \sigma_s) + \rho(v_b - v_s, \sigma_s) \} \]

<table>
<thead>
<tr>
<th></th>
<th>Proposed</th>
<th>Black-Anandan 96</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brightness constr</td>
<td>Matching-based</td>
<td>Gradient-based</td>
</tr>
<tr>
<td>Scales ( \sigma_s, \sigma_s )</td>
<td>Local adaptive</td>
<td>Rigid+tuning</td>
</tr>
<tr>
<td>Contral para</td>
<td>Constant</td>
<td>Tuning</td>
</tr>
</tbody>
</table>

Solution Technique

- Large-scale nonconvex problem
  - Statistical relaxation: slow
  - Graduated NonConvexity: LS initialization, scales control annealing
- **Our strategy**
  - Fastest descent
  - 3-step graduated optimization
  - Two sub-optimal formulations
    - Provide robust initial estimates
    - Gradually learn the local parameters
I: OFC-Based Local Regression

- Lucas-Kanade constraint: $AV = b$
- High-breakdown criterion (LMS/LTS)
- Fast deterministic algorithm
  - Least-squares (LS) initial estimate
  - Propagate using an LMS-LS procedure
  - Adaptive outlier resistance
  - Faster, more stable accuracy
- Estimate scales $\sigma_y, \sigma_z$ from inliers

II: OFC-Based Global Optimization

- **Given** $V, \sigma_y, \sigma_z$, **find** $\Delta V$ to minimize
  
  $E(\Delta V) = \sum_{i \in \text{sites}} \{p(e_\delta(\Delta V_i), \sigma_y) + \frac{1}{8} \sum_{n=1}^{N_v} p(|V_i + \Delta V_i - V_i - \Delta V_i |, \sigma_z)\}$

- **Solution:** Successive Over Relaxation
  
  $u_{\text{new}} = u_{\text{old}} - \frac{\partial E}{\partial u_{\text{old}}}, \quad T(\mu) = I_x^2 + \frac{8}{\sigma_y^2}$

  - Adaptive step size
  - Initial has dominantly high-freq errors
  - Fast convergence

III: Minimizing the Global Energy

- **Given** $V_{\text{initial}}$
- Calculate $\sigma_y, \sigma_z$
- Fastest descent by propagation
  
  - Generate candidates: $V_i \in \{V, i \in N_v; \overline{V}_i\}$
  - Replace $V_i$ by $\overline{V}_i$ if global energy $E$ drops

Hierarchical Process

- Handle large motions (>2 pixels/frame)
- Limitations:
  - Sub-sampling, warping and projection errors
  - May become the accuracy bottleneck
- Step III directly works on the image data and is less sensitive to such errors
**Overall Algorithm**

- **Image pyramid**
  - $I^0$
  - $I^1$
  - $I^{p-1}$
  - $I^p$

- **Level p**
  - $V^p_0$
  - $V^p_w$
  - $V^p_1$
  - $V^p_s$

- **Calculate gradients**
  - $V^p_0_{Iw}$

- **Local OFC**
  - $\Delta V^p_{Iw}, \sigma_{Iw}$

- **Global OFC**
  - $\Delta V^p_{Iw}$

- **Global matching**
  - $\nabla V^p_{Iw}$

- **Projection**

**Advantages**

- **Best of Everything**
  - Local OFC
    - High-quality initial flow estimates
    - Robust local scale estimates
  - Global OFC
    - Improve flow smoothness
  - Global Matching
    - The optimal formulation
    - Correct errors caused by poor gradient quality and hierarchical process

- **Results**: fast convergence, high accuracy, simultaneous motion boundary detection

**Experiments**

**Quantitative Measures**

- **True**: $V_0 = (u_0, v_0)'$, **estimate** $V = (u, v)'$

- **Our error measure**
  $$ e = \left( |u-u_0|, |v-v_0| \right)_{all sites} $$

- **Cdf curve of** $e$, **Average**: $\overline{e}$

- **Barron’s angular error** [Barron 94]
  $$ e_2 = \overline{\Delta \theta}, \Delta \theta = \arccos\left( \frac{(V^p_0,1) \cdot (V^p,1)^T}{|V^p_0,1| \cdot |V^p,1|} \right) $$

- **Error magnitude**:
  $$ e_1 = |\overline{\Delta V}|_{\text{pixels}}, |\Delta V| = |V - V_0| $$
TS: Translating Squares

- Homebrew, ideal setting, test performance
- upper bound

64x64, 1 pixel/frame

Groundtruth (cropped),
Our estimate looks the same

TS: Flow Estimate Plots

LS  BA  S1 (S2 is close)

S3 looks the same as the groundtruth.

- S1, S2, S3: results from our Step I, II, III (final)

TS: Quantitative Comparison

<table>
<thead>
<tr>
<th></th>
<th>$e_x(\cdot)$</th>
<th>$e_y(\cdot)$</th>
<th>$\pi(\cdot)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>6.14</td>
<td>0.151</td>
<td>0.0925</td>
</tr>
<tr>
<td>BA</td>
<td>8.04</td>
<td>0.209</td>
<td>0.120</td>
</tr>
<tr>
<td>BA'</td>
<td>5.88</td>
<td>0.149</td>
<td>0.0815</td>
</tr>
<tr>
<td>S1</td>
<td>1.09</td>
<td>0.0286</td>
<td>0.0180</td>
</tr>
<tr>
<td>S2</td>
<td>1.09</td>
<td>0.0284</td>
<td>0.0179</td>
</tr>
<tr>
<td>S3</td>
<td>1.15e-2</td>
<td>3.50e-4</td>
<td>2.23e-4</td>
</tr>
</tbody>
</table>

TT: Translating Tree

150x150 (Barron 94)

<table>
<thead>
<tr>
<th></th>
<th>$e_x(\cdot)$</th>
<th>$e_y(\cdot)$</th>
<th>$\pi(\cdot)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>2.60</td>
<td>0.128</td>
<td>0.0724</td>
</tr>
<tr>
<td>S3</td>
<td>0.248</td>
<td>0.0167</td>
<td>0.00984</td>
</tr>
</tbody>
</table>
**DT: Diverging Tree**

150x150 (Barron 94)

<table>
<thead>
<tr>
<th></th>
<th>ε_ρ ((\sigma))</th>
<th>ε_p ((\mu))</th>
<th>ε_{\bar{\rho}} ((\mu))</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>6.36</td>
<td>0.182</td>
<td>0.114</td>
</tr>
<tr>
<td>S3</td>
<td>2.60</td>
<td>0.0813</td>
<td>0.0507</td>
</tr>
</tbody>
</table>

**DTTT: Motion Discontinuities**

TT + DT + cookie-cutters

<table>
<thead>
<tr>
<th></th>
<th>ε_ρ ((\sigma))</th>
<th>ε_p ((\mu))</th>
<th>ε_{\bar{\rho}} ((\mu))</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>10.9</td>
<td>0.331</td>
<td>0.204</td>
</tr>
<tr>
<td>S3</td>
<td>4.03</td>
<td>0.132</td>
<td>0.0807</td>
</tr>
</tbody>
</table>

**YOS: Yosemite Fly-Through**

316x252 (Barron, cloud excluded)

<table>
<thead>
<tr>
<th></th>
<th>ε_ρ ((\sigma))</th>
<th>ε_p ((\mu))</th>
<th>ε_{\bar{\rho}} ((\mu))</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>2.71</td>
<td>0.185</td>
<td>0.118</td>
</tr>
<tr>
<td>S3</td>
<td>1.92</td>
<td>0.120</td>
<td>0.0776</td>
</tr>
</tbody>
</table>

**DTTT**

- \(u, v\)-components as intensity images
- Ours:
- BA:

```
```

```
**TAXI: Hamburg Taxi**

- Image dimensions: 256x190
- Max speed: 3.0 pix/frame
- LMS
- BA

**Pepsi Can**

- Image dimensions: 201x201 (Black)
- Max speed: 2 pix/frame
- BA

**Traffic**

- Image dimensions: 512x512 (Nagel)
- Max speed: 6.0 pix/frame

**FG: Flower Garden**

- Image dimensions: 360x240 (Black)
- Max speed: 7 pix/frame
- BA
- LMS

Ours
Error map
Smoothness error
Conclusion and Discussion

Contributions (1/2)

- Formulation
  - More complete design, minimal parameter tuning
    - Adaptive local scales
    - Strength of two error terms automatically balanced
  - 3-frame matching to avoid visibility problems
- Solution: 3-step optimization
  - Robust initial estimates and scales
  - Model parameter self-learning
  - Inherit merits of 3 methods and overcome shortcomings

Contributions (2/2)

- Results
  - High accuracy
  - Fast convergence
  - By product: motion boundaries
- Significance
  - Foundation for higher-level (model-based) visual motion analysis
  - Methodology applicable to other low-level vision problems

Future Work

- Applications
  - Non-rigid motion estimation (medical, human)
  - Higher-level visual motion analysis
    - Motion segmentation, model selection
    - Occlusion reasoning
    - Layered / contour-based representation
  - Warping w/ discontinuities
- Refinement
  - Bayesian belief propagation (BBP)
  - Better global optimization (BBP, Graph cuts etc)
A Motion-Based Bayesian Approach to Aerial Point-Target Detection and Tracking

The Problem

- UAV See And Avoid System
- Point target detection and tracking

Motion-Based Bayesian Detection

- **Background motion:**
  - Parametric optical flow

- **Object candidates:**
  - Fitting outliers
  - Motion: 3x3 SSD + fitting

- **Independent motion**
  - $\chi^2$ test

- **Bayesian mode**
  - Augment candidate set
  - Validate/update motion

Motion-Based Bayesian Detection

- **State variable:** 2D position and velocity
- **Track initialization, termination and maintenance**

The Algorithm

- Image sequence
- Motion-Based Bayesian Detection
- Measurement & Covariance
- Kalman Filter Tracking
- State & Covariance

- Prediction
- Prior
Experiments

- **1800-frame data:**
  - One target 1x2-3x3
  - Clutter (ground objects)
  - Camera wobbling
  - Low image quality

- **Results**
  - Target in track since 2nd frame
  - No false detection
  - Error: $\text{mean}=0.88$, $\text{sd}=0.44$ pixels

- Show demo

Publications


- Book Chapter

- Submission/Preparation

- Conference Papers