Is Seeing Really Believing?

A Probabilistic Framework for Confident Optical Flow Estimation

Researcher: Andy Wang, Class of 2027
Advisor: Dr. Mario Micheli, Senior Lecturer
Department of Applied Mathematics & Statistics

Johns Hopkins Whiting School of Engineering

Introduction

How does a robotic medical assistant maintain a steady view of a pulsating artery? How does a self-driving car detect a pedestrian stepping into its path? The answer lies in **optical flow** - apparent motion of objects in video.

But classical algorithms often treat motion as certain, leaving its inherent chaos and uncertainty unaddressed, and critical questions unanswered:

How much **variation** in the robot's view is introduced by light reflections?

Is the car **confident** enough to brake or swerve in the middle of the road?

There is much less literature exploring probabilistic optical flow computation, especially w.r.t. interpretable methods.

Objectives

- Compute a dense optical flow field: motion of every pixel in an image pair.
- Compute it **probabilistically**: propose a distribution for motion at each pixel, then compute its mean and variance.
- Rely on prior motion distributions when image information fails to provide a strong motion signal.
- Account for motions at varying scales: propagate distributions across image resolution levels.
- Show motion uncertainty correlates with image structure, enabling downstream applications to use confidence information.
- Run quickly: allow for real-time use.
- Remain interpretable: avoid black-box methods and training on ground-truth.

Data & Eval

Dataset

- MPI-Sintel (2012)
- 436 x 1024 image resolution
- 23 image sequences, 1041 pairs
- Albedo: obeys brightness constancy
- Clean: natural shading & reflections
- Final: blur & atmospheric effects

Baselines

- Pyramidal-LK: assume constant motion in small windows, estimate using local image intensity change.
- Farneback: approximate image intensity as quadratic polynomials, estimate by comparing coefficients.

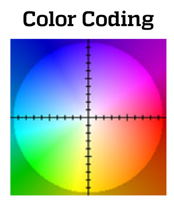
Evaluation

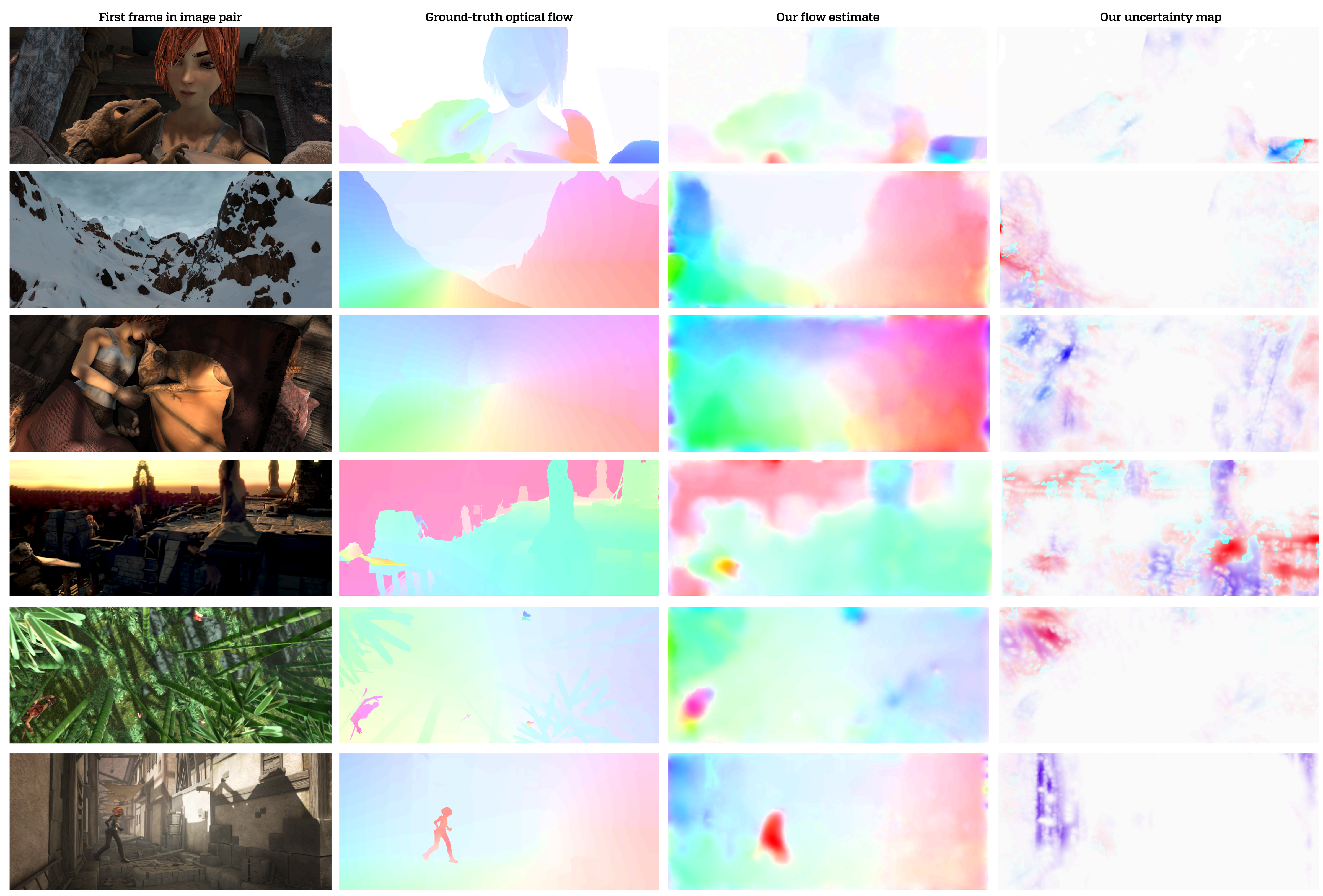
- Endpoint Error (EPE): Euclidean distance between true and estimated flow vector averaged over all pixels.

Results

1

Color Coding





2

THE INNOVATIONS

- To capture motions at varying scales, we employ a Gaussian image pyramid, performing multiple estimations at different resolution levels.
- We design a Kalman filter, a mathematical system which allows us to propagate estimates and uncertainties down our pyramid, at each level refining our prior prediction using image detail reintroduced by a higher resolution.
- The updated estimate is a weighted average of its prediction and this new data term, as is the updated uncertainty respectively, providing a fallback in case image regions contain insufficient information.

Performance	EPE (Clean)	EPE (Final)	Avg Time (s)
Pyramidal-LK	13.976	19.540	0.884
Farneback	11.369	11.523	0.326
Ours	10.509	10.832	1.010

Improvement	EPE (Clean)	EPE (Final)
Pyramidal-LK	24.8%	44.6%
Farneback	3.3%	6.0%

EPE Increase	Pyramidal-LK	Farneback	Ours
Final vs Clean	39.8%	6.0%	3.1%

3

THE MATH: optional :)

Pyramid Construction

$$I^{(\ell+1)} = 4G * \text{downsample}\left(I^{(\ell)}\right)$$

Kalman Filter

$$\mathbf{f}^{(\ell)} = 2\,\text{upsample}\left(\mathbf{f}^{(\ell+1)}\right) + \mathbf{w}^{(\ell)}$$

$$\mathbf{0} = \left[I_2\left(\mathbf{x}_i + \mathbf{f}^{(\ell)}\right) - I_1\left(\mathbf{x}_i\right)\right]_{i \in \mathcal{W}} + \mathbf{v}^{(\ell)}$$

State Transition

$$\mathbf{f}^{(\ell|\ell+1)} = 2\,\text{upsample}\left(\mathbf{f}^{(\ell+1|\ell+1)}\right)$$

$$\mathbf{\Sigma}^{(\ell|\ell+1)} = 4\,\text{upsample}\left(\mathbf{\Sigma}^{(\ell+1|\ell+1)}\right) + \mathbf{Q}^{(\ell)}$$

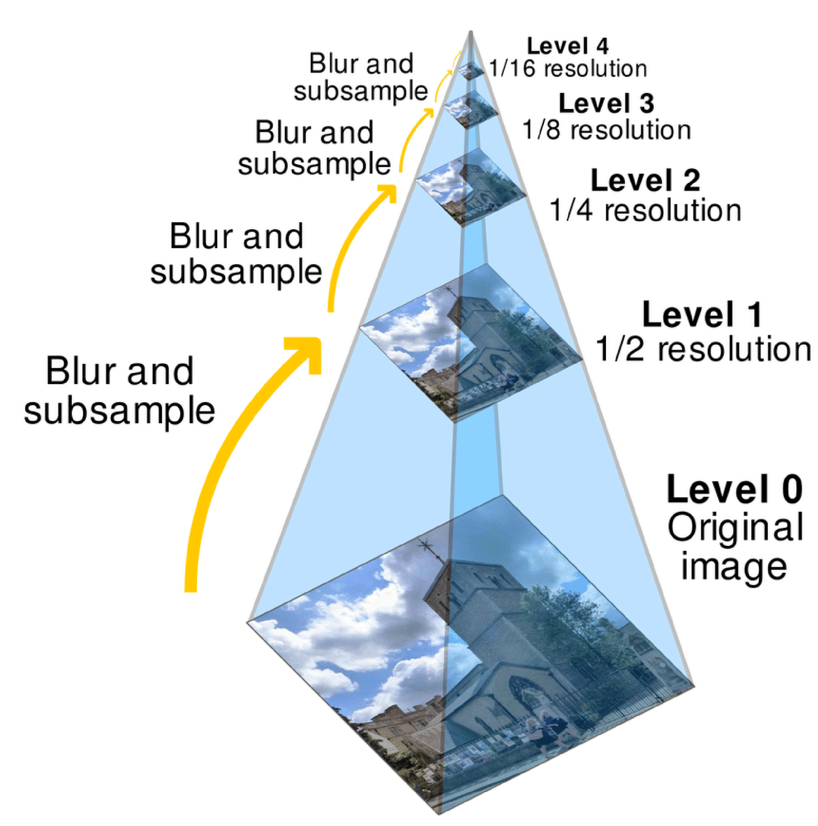
Measurement Update

$$\mathbf{Y}_{(k)} = \left(\mathbf{\Sigma}^{(\ell|\ell+1)}\right)^{-1} + \mathbf{H}_{(k)}^T \mathbf{R}_{(k)}^{-1} \mathbf{H}_{(k)}$$

$$\mathbf{y}_{(k)} = \left(\mathbf{\Sigma}^{(\ell|\ell+1)}\right)^{-1} \mathbf{f}^{(\ell|\ell+1)} + \mathbf{H}_{(k)}^T \mathbf{R}_{(k)}^{-1} \mathbf{z}_{(k)}$$

$$\mathbf{f}^{(\ell|\ell)} = \mathbf{Y}_{(K)}^{-1} \mathbf{y}_{(K)} \qquad \mathbf{\Sigma}^{(\ell|\ell)} = \mathbf{Y}_{(K)}^{-1}$$

Gaussian Image Pyramid



Flow & uncertainty are initialized to 0 & ∞. Each level's estimate is its previous level's estimate, refined based on its new image information. Level 0 produces final output.

Conclusion

Key Takeaways

- We achieve **lower EPE** than Pyramidal-LK and Farneback on MPI-Sintel.
- We output **covariance maps** showing every motion estimate's reliability.
- We provide a **fast interpretable** alternative to black-box methods.

Next Steps

- Enforce motion smoothness via a more complex state transition model.
- Experiment with different estimations of measurement noise covariance.
- Extend estimate and uncertainty propagation into temporal dimension.