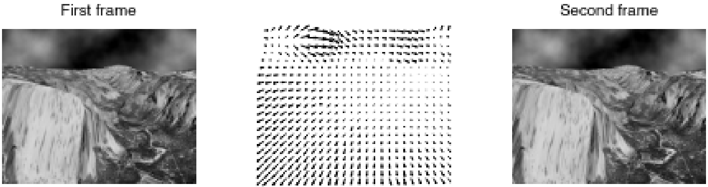
Spatial Gaussian Process Regression for Probabilistic Optical Flow Estimation

Andy Wang

8 May 2025

Table of Contents

Motion in Image Sequences



$I(x, y, t)$ $\mathbf{f}(x, y, t)$ $I(x, y, t + \delta t)$

Given two sequential images $I : \mathcal{D} \times [0, T] \to \mathcal{R}$, i.e.

$$I(x, y, t) \text{ and } I(x, y, t + \delta t)$$

(for fixed t and δt), the optical flow at time t is a vector field

$$\mathbf{f}(x, y, t) = (u(x, y, t), v(x, y, t))$$

that transforms one image into the next:

$$I\left(x + u(x, y, t), y + v(x, y, t), t + \delta t\right) \simeq I(x, y, t)$$

Barron, J. L., Fleet, D. J., & Beauchemin, S. S. (1994). Performance of Optical Flow Techniques. *International Journal of Computer Vision, 12*(1), 43-77.

Horn-Schunck Optical Flow Computation

Data correlation constraint: object brightness remains constant while location changes.

$$\frac{d}{dt}I(x(t), y(t), t) = \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0, \quad u := \frac{dx}{dt}, v := \frac{dy}{dt}$$

Its solution is not unique, so we introduce regularity of u, v, i.e. small $\nabla u, \nabla v$.
We determine u, v for each x, y by minimizing an energy (loss) function:

$$E[u, v] = \iint_D \left[\left(\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} \right)^2 + \alpha \left(\|\nabla u\|^2 + \|\nabla v\|^2 \right) \right] dx\, dy$$

Using calculus of variations, we obtain the Euler-Lagrange equations:

$$\frac{\partial I}{\partial x}\left(\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} \right) - \alpha \nabla^2 u = 0 \qquad \frac{\partial I}{\partial y}\left(\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} \right) - \alpha \nabla^2 v = 0$$

Approximating $\nabla^2 u \approx \bar{u} - u$, we obtain an iterative scheme:

$$u^{k+1} = \bar{u}^k - I_x \left(\frac{I_x \bar{u}^k + I_y \bar{v}^k + I_t}{\alpha^2 + I_x^2 + I_y^2} \right), \quad v^{k+1} = \bar{v}^k - I_y \left(\frac{I_x \bar{u}^k + I_y \bar{v}^k + I_t}{\alpha^2 + I_x^2 + I_y^2} \right)$$

Horn, B. K. P., & Schunck, B. G. (1981). Determining Optical Flow. *Artificial Intelligence, 17*(1-3), 185-203.

Lucas-Kanade Optical Flow Computation

Same assumption (brightness constancy) as Horn-Schunck:

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0$$

We assume same $(u(x,y), v(x,y))\ \forall (x,y) \in W$. The constraints for W become:

$$\begin{bmatrix} I_x(x_1,y_1) & I_y(x_1,y_1) \\ \vdots & \vdots \\ I_x(x_n,y_n) & I_y(x_n,y_n) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -I_t(x_1,y_1) \\ \vdots \\ -I_t(x_n,y_n) \end{bmatrix}$$

This system is overdetermined, so its least squares solution is $\mathbf{v} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}$:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum I_x(x_i,y_1)^2 & \sum I_x(x_i,y_i)I_y(x_i,y_i) \\ \sum I_x(x_i,y_i)I_y(x_i,y_i) & \sum I_y(x_i,y_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum I_x(x_i,y_i)I_t(x_i,y_i) \\ -\sum I_y(x_i,y_i)I_t(x_i,y_i) \end{bmatrix}$$

We may introduce weights via diagonal matrix $\mathbf{W}$ to prioritize certain pixels, yielding:

$$\mathbf{v} = \left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}\mathbf{b}$$

Lucas, B. D., & Kanade, T. (1981). An Iterative Image Registration Technique with an Application to Stereo Vision. *7th International Joint Conference on Artificial Intelligence*, *2*, 674-679.

Probabilistic Motivation

Classical methods produce deterministic point estimates.
Uncertainty is inherent in optical flow estimation, due to:

- Image noise
- Brightness changes
- Low contrast regions
- Object occlusion
- Aperture problem
- Incompatible motions in localized regions

Quantification of confidence is desirable for critical aplications, such as:

- Autonomous navigation
- Computer-integrated surgery/healthcare
- Real-time surveillance

1D Gaussian Processes

A Gaussian process is a infinite collection of random variables:

$$\{X_t\}_{t\in T} \sim \mathcal{GP}(m(t), k(t,t'))$$

such that every finite subset of $\{X_t\}_{t\in T}$ is multivariate Gaussian:

$$\begin{bmatrix} X_{k_1} \\ \ddots \\ X_{k_n} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(t_{k_1}) \\ \vdots \\ m(t_{k_n}) \end{bmatrix}, \begin{bmatrix} k(t_{k_1}, t_{k_1}) & \cdots & k(t_{k_1}, t_{k_n}) \\ \vdots & \ddots & \vdots \\ k(t_{k_n}, t_{k_1}) & \cdots & k(t_{k_n}, t_{k_n}) \end{bmatrix}\right)$$

Common mean functions include:

$$m(t) = 0,\ \ m(t) = \mu,\ \ m(t) = at + b,\ \ m(t) = a_0 + a_1 t + \cdots + a_n t^n$$

Common covariance kernels include:

$$k(t,t') = \sigma^2 \exp\left(-\frac{(t-t')^2}{2\lambda^2}\right),\ k(t,t') = \sigma^2 \exp\left(-\frac{|t-t'|}{\lambda}\right)$$

Mean/covariance encode process assumptions.

2D Gaussian Processes

Spatial Gaussian process:

$$\{\mathbf{f}(\mathbf{x})\}_{\mathbf{x}\in\mathcal{D}} = \left\{ \begin{bmatrix} \mathbf{u}(\mathbf{x}) \\ \mathbf{v}(\mathbf{x}) \end{bmatrix} \right\}_{\mathbf{x}\in\mathcal{D}} \sim \mathcal{GP}(\mathbf{m}(\mathbf{x}), \mathbf{k}(\mathbf{x}, \mathbf{x}'))$$

Example mean/covariance:

$$\mathbf{m}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}, \ \ \mathbf{k}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{(t-t')^2}{2\lambda^2} \right) \boldsymbol{\Sigma}$$

Their joint distribution is

$$\begin{bmatrix} \mathbf{f}(\mathbf{x}_1) \\ \vdots \\ \mathbf{f}(\mathbf{x}_n) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}(\mathbf{x}_1) \\ \ddots \\ \mathbf{m}(\mathbf{x}_n) \end{bmatrix}, \begin{bmatrix} \mathbf{k}(\mathbf{x}_1, \mathbf{x}_1) & \cdots & \mathbf{k}(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ \mathbf{k}(\mathbf{x}_n, \mathbf{x}_1) & \cdots & \mathbf{k}(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix} \right)$$

A reformulation for easier interpretability (2D):

$$\mathbf{F} := \begin{bmatrix} \mathbf{U}(\mathbf{X}) \\ \mathbf{V}(\mathbf{X}) \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{\mu} := \begin{bmatrix} \boldsymbol{\mu}_u \\ \boldsymbol{\mu}_v \end{bmatrix}, \mathbf{K} := \begin{bmatrix} \mathbf{K}_{uu} & \mathbf{K}_{uv} \\ \mathbf{K}_{vu} & \mathbf{K}_{vv} \end{bmatrix} \right)$$

Gaussian Process Regression

Through deterministic methods, we obtain noisy observations $\tilde{\mathbf{F}}$:

$$\tilde{\mathbf{f}}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) + \boldsymbol{\eta}(\mathbf{x}),\ \ \boldsymbol{\eta}(\mathbf{x}) \sim \mathcal{N}\left(\mathbf{0}, \sigma_\eta^2 \mathbf{I}\right)$$

The noisy distribution is:

$$\tilde{\mathbf{F}} \big| \mathbf{F} \sim \mathcal{N}\left(\mathbf{F}, \sigma_\eta^2 \mathbf{I}\right),\ \ \mathbf{F} \big| \mathbf{X} \sim \mathcal{N}\left(\boldsymbol{\mu}, \mathbf{K}\right) = \tilde{\mathbf{F}} \implies \tilde{\mathbf{F}} \big| \mathbf{X} \sim \mathcal{N}\left(\boldsymbol{\mu}, \mathbf{K} + \sigma_\eta^2 \mathbf{I}\right)$$

We want to estimate true optical flows: we form the joint distribution.

$$\begin{bmatrix} \mathbf{F} \\ \tilde{\mathbf{F}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \mathbf{K} \\ \mathbf{K} & \mathbf{K} + \sigma_\eta^2 \mathbf{I} \end{bmatrix} \right)$$

We derive the conditional distribution:

$$\mathbf{F} | \tilde{\mathbf{F}} \sim \mathcal{N}\left(\boldsymbol{\mu} + \mathbf{K}\left(\mathbf{K} + \sigma_\eta^2 \mathbf{I}\right)^{-1} \left(\tilde{\mathbf{F}} - \boldsymbol{\mu}\right), \mathbf{K} - \mathbf{K}\left(\mathbf{K} + \sigma_\eta^2 \mathbf{I}\right)^{-1} \mathbf{K} \right)$$

Mean serves as estimate, covariance serves as uncertainty.

Gaussian Process Regression (cont.)

We obtain incomplete observations $\mathbf{F}_X$ and want to interpolate $\mathbf{F}_*$:

$$\begin{bmatrix} \mathbf{F}_* \\ \mathbf{F}_X \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_* \\ \boldsymbol{\mu}_X \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{**} & \mathbf{K}_{*X} \\ \mathbf{K}_{X*} & \mathbf{K}_{XX} \end{bmatrix} \right)$$

The conditional distribution is reformulated:

$$\mathbf{F}_* | \mathbf{F}_X \sim \mathcal{N}\left(\boldsymbol{\mu}_* + \mathbf{K}_{*X} \left(\mathbf{K}_{XX} + \sigma_\eta^2 \mathbf{I}\right)^{-1} (\mathbf{F}_X - \boldsymbol{\mu}_X), \mathbf{K}_{**} - \mathbf{K}_{*X} \left(\mathbf{K}_{XX} + \sigma_\eta^2 \mathbf{I}\right)^{-1} \mathbf{K}_{X*} \right)$$

Parameter Fitting

Mean/covariance parameters (e.g. $\sigma^2, \lambda^2, \sigma_\eta^2$) are optimized by maximizing the likelihood of observing $\tilde{\mathbf{F}}$. This likelihood is:

$$p\left(\tilde{\mathbf{F}}|\mathbf{X}\right) \propto \frac{1}{\sqrt{\det\left(\mathbf{K}+\sigma_\eta^2\mathbf{I}\right)}} \exp\left(-\frac{1}{2}\left(\tilde{\mathbf{F}}-\boldsymbol{\mu}\right)^T \left(\mathbf{K}+\sigma_\eta^2\mathbf{I}\right)^{-1}\left(\tilde{\mathbf{F}}-\boldsymbol{\mu}\right)\right)$$

We use the negative log-likelihood for convenience:

$$-\log p\left(\tilde{\mathbf{F}}|\mathbf{X}\right) \propto \frac{1}{2}\left(\tilde{\mathbf{F}}-\boldsymbol{\mu}\right)^T \left(\mathbf{K}+\sigma_\eta^2\mathbf{I}\right)^{-1}\left(\tilde{\mathbf{F}}-\boldsymbol{\mu}\right) + \frac{1}{2}\det\left(\mathbf{K}+\sigma_\eta^2\mathbf{I}\right)$$

We minimize this numerically via gradient descent.

Parameter Fitting (cont.)

Let φ be a parameter of $\mathbf{m}$, $\mathbf{y} = \tilde{\mathbf{F}} - \boldsymbol{\mu}$, $\mathbf{K}_y = \mathbf{K} + \sigma_\eta^2 \mathbf{I}$:

$$\frac{\partial \mathcal{L}}{\partial \varphi} = \frac{1}{2}\left[\left(\frac{\partial \mathbf{y}^T}{\partial \varphi}\right)\mathbf{K}_y^{-1}\mathbf{y} + \mathbf{y}^T\mathbf{K}_y\left(\frac{\partial \mathbf{y}}{\partial \varphi}\right)\right] = \frac{1}{2}\left(2\mathbf{y}^T\mathbf{K}_y\left(\frac{\partial \mathbf{y}}{\partial \varphi}\right)\right) = -\mathbf{y}^T\mathbf{K}_y\left(\frac{\partial \mathbf{m}}{\partial \varphi}\right)$$

Let θ be a parameter of $\mathbf{k}$:

$$\frac{\partial \mathcal{L}}{\partial \theta} = \mathbf{y}^T\left(\frac{\partial \mathbf{K}_y^{-1}}{\partial \theta}\right)\mathbf{y} + \frac{\partial |\mathbf{K}_y|}{\partial \theta} = -\frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\left(\frac{\partial \mathbf{K}_y}{\partial \theta}\right)\mathbf{K}^{-1}\mathbf{y} + \frac{1}{2}\mathrm{tr}\left(\mathbf{K}_y^{-1}\left(\frac{\partial \mathbf{K}_y}{\partial \theta}\right)\right)$$

RBF kernel:

$$\frac{\partial \mathbf{K}_y}{\partial \sigma} = \frac{2\mathbf{K}}{\sigma},\ \frac{\partial \mathbf{K}_y}{\partial \lambda} = \mathbf{K} \odot \frac{(\mathbf{x} - \mathbf{x}')}{\lambda^3},\ \frac{\partial \mathbf{K}_y}{\partial \sigma_\eta} = 2\sigma_\eta \mathbf{I}$$

Implementation

Code implementations of all algorithms were written in Python.

The Yosemite sequence consists of 15 252×316 RGBA images, first grayscaled.

Spatial derivatives were computed using Sobel filters.

Temporal derivatives were computed using simple differencing.

Horn-Schunck was performed for 100 iterations using $\lambda = 500$.

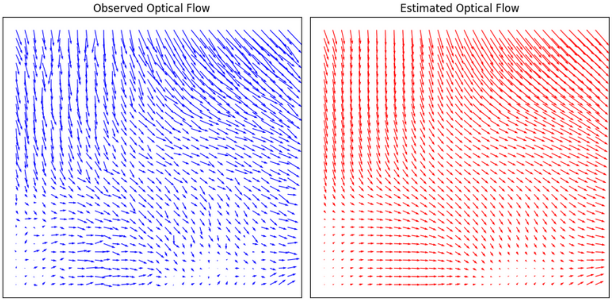
Lucas-Kanade was performed with 35×35 windows.

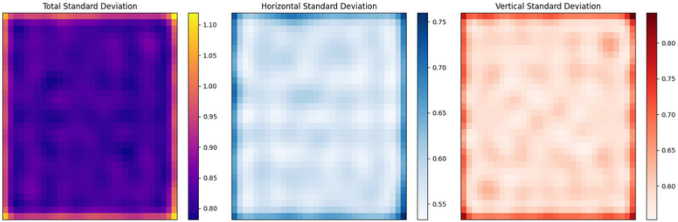
Gaussian processes were created/fit using PyTorch.

GPs were trained on 25×31 downsampled flow fields for memory purposes.

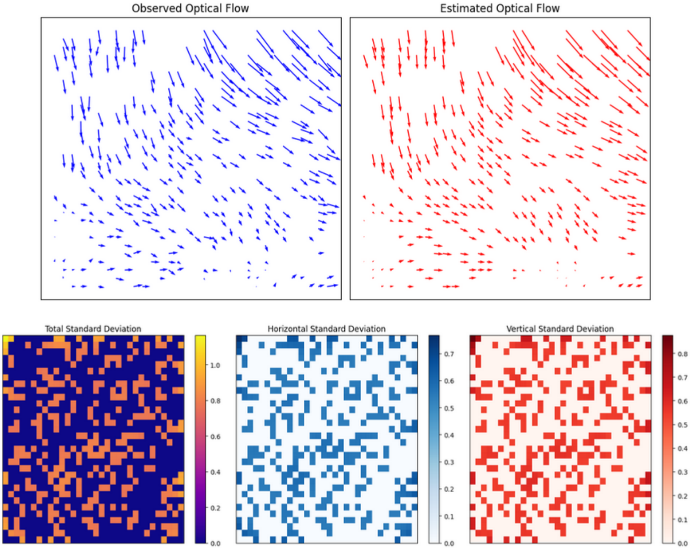
Training ran for 100 iterations using Adam with learning rate $\eta = 0.1$.

Results (Synthetic, Smoothing)

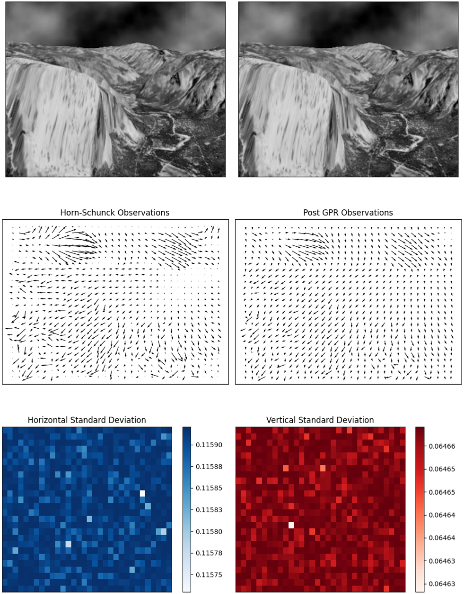




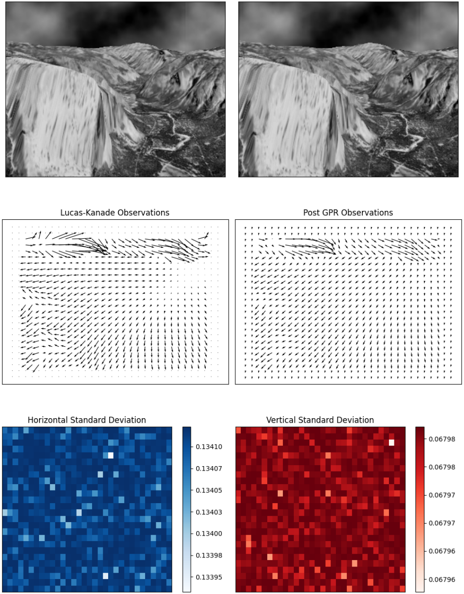
Results (Synthetic, Interpolation)



Results (Yosemite, Horn-Schunck)



Results (Yosemite, Lucas-Kanade)



Next Steps

We've successfully enforced spatial smoothness and quantified confidence.

- Evaluate against ground truth:

$$\text{EPE} = \frac{1}{N} \sum_{i=1}^{N} \sqrt{(u_i - \hat{u}_i)^2 + (v_i - \hat{v}_i)^2}$$

- Obtain better observations: FlowNet, RAFT.
- Experiment with more complex datasets: KITTI, MPI Sintel, FlyingChairs.
- Compare different hyperparameters, identify optimal conditions:
  Ex: when should I use L_1 versus L_2 covariance?
- Capture nonhomogeneous noise
- Exploit efficient representations of sparse flows
- Perform segmentation, smooth each segment independently
- Extend to 3D spatiotemporal regression (3+ frames)

Ideas? Questions? Comments?