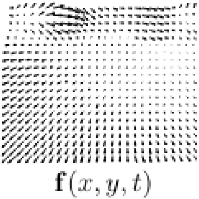
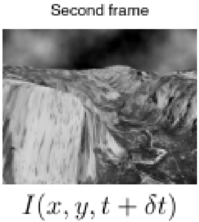
Spatial Gaussian Process Regression for Bayesian Optical Flow Estimation

Andy Wang & Luca Orquiza

2 December 2025

Motion in Image Sequences





Given two sequential images $I : \mathcal{D} \times [0, T] \rightarrow \mathcal{R}$, 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

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

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.

Lucas-Kanade Method

Brightness constancy assumption:

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

Underdetermined, so assume same $(u(x,y), v(x,y))$ $\forall (x,y) \in W$:

$$\underbrace{\begin{bmatrix} I_x(x_1, y_1) & I_y(x_1, y_1) \\ \vdots & \vdots \\ I_x(x_n, y_n) & I_y(x_n, y_n) \end{bmatrix}}_{\boldsymbol{\nabla I}^T} \underbrace{\begin{bmatrix} u \\ v \end{bmatrix}}_{\boldsymbol{f}} = \underbrace{\begin{bmatrix} -I_t(x_1, y_1) \\ \vdots \\ -I_t(x_n, y_n) \end{bmatrix}}_{-\boldsymbol{I_t}}$$

Overdetermined: least squares solution is $\boldsymbol{v} = \left(\boldsymbol{A}^T \boldsymbol{A}\right)^{-1} \boldsymbol{A}^T \boldsymbol{b}$:

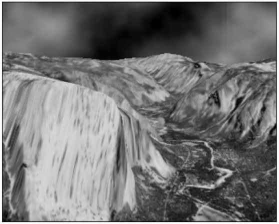
$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum I_x(x_i, y_i)^2 & \sum I_x(x_i, y_i) I_y(x_i, y_i) \\ \sum I_x(x_i, y_i) I_y(x_i, y_i) & \sum I_y(x_i, y_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum I_x(x_i, y_i) I_t(x_i, y_i) \\ -\sum I_y(x_i, y_i) I_t(x_i, y_i) \end{bmatrix}$$

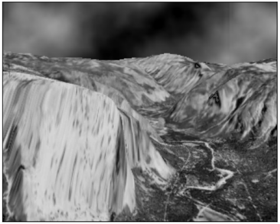
Optional weighting of pixels via diagonal matrix $\boldsymbol{W}$:

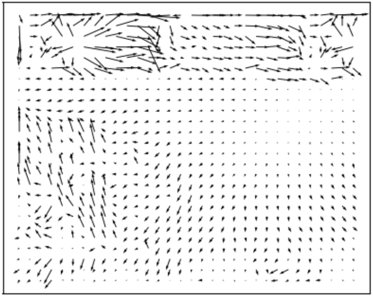
$$\boldsymbol{v} = \left(\boldsymbol{A}^T \boldsymbol{W} \boldsymbol{A}\right)^{-1} \boldsymbol{A}^T \boldsymbol{W} \boldsymbol{b}$$

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.

Lucas-Kanade Results







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

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 safety-critical aplications, e.g.:

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

Review of Flow Methods

Existing deterministic optical flow methods include:

- Lucas & Kanade (1981): previously described.
- Horn & Schunck (1981): estimate optical flow by minimizing a global energy functional (with smoothness regularizer) via calculus of variations.
- FlowNet (Dosovitskiy et al., 2015) use an encoder-decoder CNN (encoder captures low-level semantic motion, decoder restores spatial resolution).

Existing probabilistic optical flow methods include:

- Simoncelli et al. (1991) model brightness constraint errors with Gaussian noise and derive a posterior flow distribution whose mean gives a regularized gradient-based flow estimate.
- Roy & Govindu (2000) formulate optical flow as a Markov Random Field labeling problem, solve it via graph cuts on angle and magnitude parameters.
- Wannenwetsch et al. (2017) perform variational inference on an energy-based model using mean-field approximation to predict optical flow and uncertainty as entropy of the variational distribution.

Wang-Orquiza Method

Suppose $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\eta}$, $\boldsymbol{A}$ fixed, $\boldsymbol{x}$ fixed unknown, $\boldsymbol{y}$ observed, $\boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H})$:

$$\hat{\boldsymbol{x}}_{\text{MLE}}(\boldsymbol{y}) = \arg\max_{\boldsymbol{x}} f_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{x}) = \arg\max_{\boldsymbol{x}} f_{\boldsymbol{\eta}}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) = \boldsymbol{\Sigma}_{\hat{\boldsymbol{x}}} \boldsymbol{A}^T \boldsymbol{H}^{-1} \boldsymbol{y}$$

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{x}}} = \text{Cov}[\hat{\boldsymbol{x}}_{\text{MLE}}(\boldsymbol{y}) - \boldsymbol{x}] = (\boldsymbol{A}^T \boldsymbol{H}^{-1} \boldsymbol{A})^{-1}$$

Same assumptions as Lucas-Kanade, plus additive Gaussian noise:

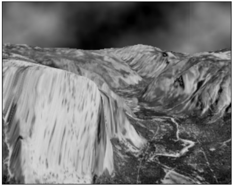
$$\boldsymbol{I_t} = -\boldsymbol{\nabla I}^T \boldsymbol{f} + \boldsymbol{\eta} \qquad \boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H})$$

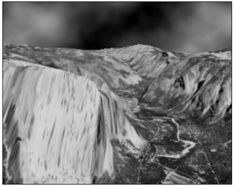
Results above give:

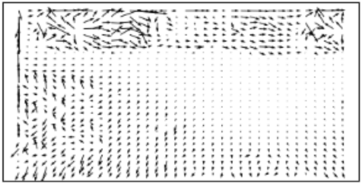
$$\hat{\boldsymbol{f}} = -\boldsymbol{\Sigma}_{\hat{\boldsymbol{f}}} \boldsymbol{\nabla I} \boldsymbol{H}^{-1} \boldsymbol{I}_t \qquad \boldsymbol{\Sigma}_{\hat{\boldsymbol{f}}} = (\boldsymbol{\nabla I} \boldsymbol{H}^{-1} \boldsymbol{\nabla I}^T)^{-1}$$

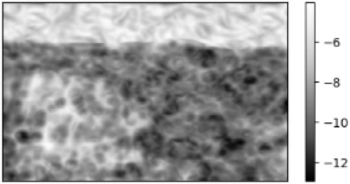
Poor texture around $\boldsymbol{x} \implies$ small $\nabla I \implies$ large $\boldsymbol{\Sigma}_{\hat{\boldsymbol{f}}}$.

Wang-Orquiza Results









Spatial Gaussian Processes

2-dimensional Gaussian process (distribution over functions of 2D inputs):

$$\boldsymbol{f}(\cdot) = \begin{bmatrix} u(\cdot) \\ v(\cdot) \end{bmatrix} \sim \mathcal{GP}(\boldsymbol{m}(\cdot), \boldsymbol{k}(\cdot, \cdot'))$$

Example mean & covariance kernel (affine & RBF):

$$\boldsymbol{m}(\cdot) = \boldsymbol{A}(\cdot) + \boldsymbol{b}, \;\; \boldsymbol{k}(\cdot, \cdot') = \exp\left(-\frac{\| \cdot - \cdot' \|^2}{2\lambda^2}\right) \boldsymbol{\Sigma}$$

Joint distribution of function values at finite subset of points:

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

Equivalently:

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

Gaussian Process Regression

Through Wang-Orquiza method, we obtain noisy observations $\tilde{\boldsymbol{F}}$:

$$\tilde{\boldsymbol{f}}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\eta}(\boldsymbol{x}),\ \boldsymbol{H}(\boldsymbol{x}) \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)$$

The noisy distribution is:

$$\tilde{\boldsymbol{F}} \big| \boldsymbol{F} \sim \mathcal{N}\left(\boldsymbol{F}, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right),\ \boldsymbol{F} \sim \mathcal{N}\left(\mu, \boldsymbol{K}\right) \implies \tilde{\boldsymbol{F}} \sim \mathcal{N}\left(\mu, \boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)$$

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

$$\begin{bmatrix} \boldsymbol{F} \\ \tilde{\boldsymbol{F}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ \mu \end{bmatrix}, \begin{bmatrix} \boldsymbol{K} & \boldsymbol{K} \\ \boldsymbol{K} & \boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}} \end{bmatrix}\right)$$

We derive the conditional distribution (posterior):

$$\boldsymbol{F} | \tilde{\boldsymbol{F}} \sim \mathcal{N}\left(\mu + \boldsymbol{K}\left(\boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)^{-1}\left(\tilde{\boldsymbol{F}} - \mu\right), \boldsymbol{K} - \boldsymbol{K}\left(\boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)^{-1}\boldsymbol{K}\right)$$

Mean serves as estimate, covariance as uncertainty.

Parameter Fitting

Mean/covariance parameters (e.g. lengthscale λ^2) are optimized by maximizing the marginal likelihood of observations $\tilde{\boldsymbol{F}}$. This likelihood is:

$$p\left(\tilde{\boldsymbol{F}}\right) \propto \frac{1}{\sqrt{\det\left(\boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)}} \exp\left(-\frac{1}{2}\left(\tilde{\boldsymbol{F}} - \boldsymbol{\mu}\right)^{T}\left(\boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\boldsymbol{I}\right)^{-1}\left(\tilde{\boldsymbol{F}} - \boldsymbol{\mu}\right)\right)$$

In practice we minimize negative log-likelihood via gradient descent:

$$-\log p\left(\tilde{\boldsymbol{F}}\right) \propto \frac{1}{2}\left(\tilde{\boldsymbol{F}} - \boldsymbol{\mu}\right)^{T}\left(\boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)^{-1}\left(\tilde{\boldsymbol{F}} - \boldsymbol{\mu}\right) + \frac{1}{2}\det\left(\boldsymbol{K} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{f}}}\right)$$

We minimize this numerically via gradient descent.

Parameter Fitting (cont.)

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

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

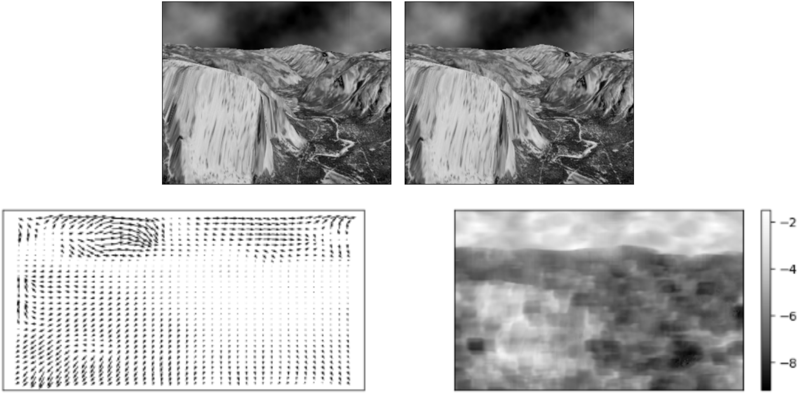
Let θ be a parameter of $\boldsymbol{k}$:

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

RBF parameter λ^2:

$$\frac{\partial \boldsymbol{K}_y}{\partial \lambda} = \boldsymbol{K} \odot \frac{\|\boldsymbol{x} - \boldsymbol{x}'\|}{\lambda^3}$$

GP Regression Results



Discussion

Our maximum-likelihood method produced an initial estimate and uncertainty.

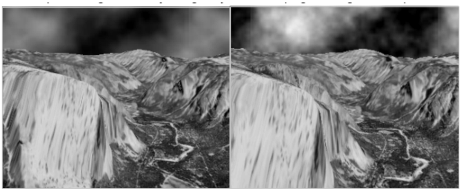
Our Gaussian-process Bayesian method provided updated these (posterior), after enforcing a global smoothness constraint (Gaussian process prior).

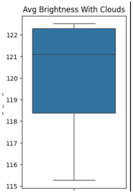
Possible next steps:

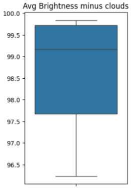
- Obtain better observations: FlowNet, RAFT, etc.
- Experiment with more complex covariance kernels.
- Perform segmentation, smooth each segment independently.
- Extend to spatiotemporal regression (3D, multiple frames).

Appendix (Regarding Our Data)

Our data for this project is the Yosemite sequence, a well-known benchmark dataset for optical flow problems. It consists of 15 simulated images of Yosemite national park “taken” by an aerial camera moving in a straight line. It was generated by Lynn Quam at SRI, by taking a topdown image and texture mapping it to a depth map to create a 3D model, with the images generated from a similar camera. Frames 1 (first) and 15 (last) make the camera’s motion more apparent. Our image data consists solely of grayscale brightness values at each pixel. No summary statistics are really helpful, and images aren’t modeled generatively, but we’ve included average brightness (with and without the images’ top “sky” section across all 15 frames):




Avg Brightness With Clouds


Avg Brightness minus clouds