Computer Vision-Optical Flow

Puranjay Datta
Electrical Engineering
IIT BOMBAY

***Abstract*—Computer vision is the field of computer science and Artificial Intelligence that deals with replicating complex functionalities of our human eye and helps computers perceive and process images/videos in the same way. Computer vision deals with various image processing tasks such as object detection, classification, segmentation, recognition. In this paper, we will be talking about Optical flow estimation. Motion perception has been an integral aspect of our visual experience. Therefore our goal is to estimate the 2D motion fields i.e 2D velocities of all visible points in the image frame. Two key problems are identifying which objects to detect and how to track them. I will be discussing the harris corner detection method to identify some corners/particles and use Lucas Kanade Algorithm to track them.**

I. Motivation

Motion is a rich source of information in conveying important aspects of any process. Optical flow is used robustly by engineers for visual odometry, movement detection, image dominant plane extraction. It not only helps in identifying the motion of an object and observer but also makes us aware of the structure of components in the environment. Consider a ball is moving to the right in a sequence of five frames. This movement is captured essentially by optical flow estimation in just one frame thus compressing the information. The contemporary area of research is biological neural circuitry and neuromorphic engineering which extensively uses optical flow sensors.

II. Introduction

Video is a combination of many frames/images and the image is a function of coordinate space(x,y) and time(t). Optical flow is defined as the apparent motion of brightness patterns in the image. Three important assumptions while calculating optical flow are
Brightness Constancy-The neighbouring points in the image are often of similar brightness assuming the surface illumination is constant.
Small motion-limited movement of particles between frames.
Spatial coherence-Neighbouring Points have similar velocities.

$$I(x, y, t-1) = I(x + u(x, y), y + v(x, y), t)\ (Assumption)$$

$$I(x+u, y+v, t) \approx I(x, y, t-1) + I_x u(x, y) + I_y v(x, y) + I_t$$

$$I(x+u, y+v, t) - I(x, y, t-1) \approx I_x u(x, y) + I_y v(x, y) + I_t$$

$$\nabla I[u\ v]^T + I_t \approx 0$$

A. Aperture Problem

The above gradient constraint provides 1 constraint in 2 unknowns u,v. The gradient constrains the velocity in the normal direction and has no control in the tangential direction.The unique Normal velocity is given as $u_n = -\frac{f_t}{\|\vec{\nabla} f\|}\frac{\vec{\nabla} f}{\|\vec{\nabla} f\|}$.
If $\vec{\nabla} f = 0$ the normal velocity is undefined hence we get no constraint.In any case, we need more constraints to find the velocity vector $\vec{u} = (u, v)^T$.This is referred to as the Aperture problem in optical flow estimation.

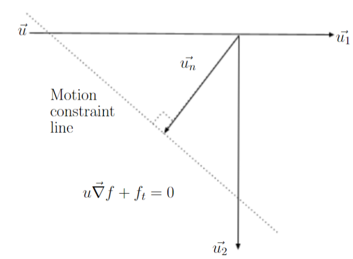


Fig. 1. Normal Velocity

III. Lucas Kanade Method(KLT)

To solve the problem of 2 unknowns and 1 constraint we provide another constraint of spatial coherence. We assume 5 5-pixel area to have the same (u,v) which gives 25 equations i.e over-constrained. There might be no velocity (u1,u2) that might satisfy all these equations hence we use the Least square estimator to minimize squared errors(LS).

$$I_t(p_i) + \nabla I(p_i).[u\ v] = 0$$

$$\begin{bmatrix} I_x(p1) & I_y(p1) \\ I_x(p2) & I_y(p2) \\ . & . \\ . & . \\ I_x(p25) & I_y(p25) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(p1) \\ I_t(p2) \\ . \\ . \\ I_t(p25) \end{bmatrix}$$

$$A\ d = b$$

Least squares solution for d is given by $(A^T A)d = A^T b$

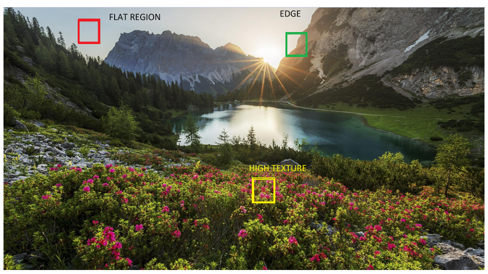
$$\underbrace{\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix}}_{A^T A} \begin{bmatrix} u \\ v \end{bmatrix} = - \underbrace{\begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix}}_{A^T b}$$

The following above equation is solvable uder the given conditions-

1) $A^T A$ should be invertible
2) Eigen values λ_1 and λ_2 should not be too small in magnitude
3) The matrix $A^T A$ should be well conditioned i.e Eigen value $\frac{\lambda_1}{\lambda_2}$ should not be too large and λ_1 being the larger of them.

The matrix $M = A^T A$ is the second-moment matrix Harris corner detector. The Eigen-vectors and Eigenvalues of $A^T A$ relate to edge direction and magnitude. The eigenvector associated with larger eigenvalue , points in the direction of the fastest intensity change and the other eigenvector is perpendicular to it.Hence Optical flow heavily depends on corner feature extraction for it's accuracy.Interpretation of eigenvalues is explained below

1) $\lambda_2 >> \lambda_1$ or $\lambda_1 >> \lambda_2 \implies$ edge.
2) $\lambda_2 \approx \lambda_1$ and are large $\implies$ corner.
3) λ_1 and λ_2 are small $\implies$ a flat region



IV. Iterative Optical Flow estimation

From the above 25 equations it is clear that KLT involves over-constrained set of equations which can be only be solved by iterative estimation.The Lucas Kanade method approximates the brightness constraint equation as a linear equation neglecting higher order terms.The estimation error is bounded by the following expression

$$|\hat{d} - d| \leq \frac{d^2 f_1''(x)}{2|f_1'(x)|} + O(d^3)$$

For sufficiently small displacements and finite $\frac{f_1''}{f_1}$ we get high accuracy.Thus we can use Gauss-Newton optimaization which useS current estimated motion to undo and then warp the signals to find new residual motion until convergence of residual motion is reached. Consider a warped 2D image transformation as follows -

$$I^0(\vec{x}, t + \delta t) = I(\vec{x} + \vec{u}^0 \delta t, t + \delta t)$$

where δt is time between consecutive frames.We have $\vec{u} = \vec{u}^0 + \delta \vec{u}$ which yields the following -

$$I^0(\vec{x}+\delta\vec{u}, t+1) = I(\vec{x}+\delta\vec{u}+\vec{u}^0, t+1) = I(\vec{x}+\vec{u}, t+1) = I^0(\vec{x}, t)$$

If $\delta \vec{u} = 0$ then we get I^0 as constant in time else we can compute the residual motion using -

$$\delta \vec{u} = M^{-1} \vec{b}$$

$$\vec{u} = \vec{u}^0 + \delta \vec{u}$$

This residual motion is used to rewarp the image and then calculate new residual motion.This sequence of iterations converges to the desired objective function and the flow estimates converge to the LS estimate.

V. Temporal Aliasing and coarse to fine Gaussian pyramids

We obtain data by sampling the continuous signals at certain rates decided by various sampling theorems. If the displacement between two consecutive frames is large (caused by the high flutter speed of the camera) temporal aliasing is caused which leads to the convergence to wrong optical flow. Sampling the image at an interval of $\frac{2\pi}{T}$ where T is the time duration between two frames, causes spectrum replicas. The derivative filters are more sensitive to spectrum replicas at high frequencies. To avoid this we use coarse to fine estimation using Gaussian pyramids.

1) Start at the base level k=L lowest resolution, warp the images and obtain LS estimated residual motion until convergence $\vec{u}_L$.
2) warp level k=L-1 by using $\vec{u}_L$ and obtain the new motion $\vec{u}_{L-1}$ until convergence.
3) ....until level k=0.

Drawback of this method is that a poor estimate at lower level leads to wrong estimate at finer levels.Aliasing may cause optimization algorithms to converge to a local minima which can be prevented by using coarse to fine approximation.

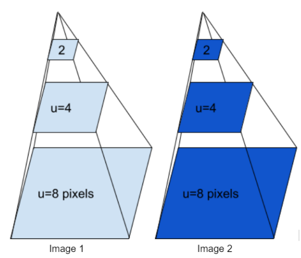


Fig. 2. Gaussian Pyramids

VI. Global smoothing using Horn Schunck Method

Unlike Lucas Kanade's method which assumes constant optical flow, Horn Schunck provides a smooth optical flow. All the objects in the world are rigid and they move coherently in a smooth fashion. Apart from the brightness constancy condition, we have another smooth flow condition. Thus the optical flow objective function is

$$E(\vec{u}) = \int (\nabla I.\vec{u} + I_t)^2 + \lambda^2 (\left\|\vec{\nabla} u_1\right\|^2 + \left\|\vec{\nabla} u_2\right\|^2)$$

The smoothness regularization coefficient is the sum of squared terms hence needs to be minimized. Therefore the texture free region is devoid of optical flow whereas on the edge the points will flow to the nearby region which is given by the aperture problem. The key benefit of global smoothing is that it helps in propagating the optical motion field to larger distances. For example, if there is a patch of uniform intensity , local optical flow methods would yield singularity whereas global methods are used to fill in the optical flow from neighbouring cells with the help of gradient constraints. The only disadvantage of global smoothing is the computational cost which is very high compared to local methods even after using good optimizing techniques. One can use precomputed gradients to speed up the process.

VII. Application of Lucas Kanade algorithm to precipitation nowcasting

The question of How much will it rain in the next hour in this area essentially refers to precipitation nowcasting. It is defined as a forecast with high spatial-temporal resolution. It helps us in predicting high convective rainfall, flash floods leading to sediment deposition and soil erosion. For every algorithm to work we need a feature to work with.

1) Identify precipitation features using shi Tomasi corner detection algorithm which uses the eigenvalues λ_1, λ_2 of the covariance matrix of Intensity derivatives and based on R score=$min(\lambda_1, \lambda_2)$ detects them.
2) Track the features at time t above by solving the set of optical flow estimation equations in the local neighbourhood as shown above in the Lucas Kanade object tracking method.
3) extrapolates the features to predict their position at lead time n.
4) Extrapolate the features at time t by warping them to their future locations using the affine transformation matrix. The remaining discontinuities are then linearly interpolated to estimate their nowcast intensities.

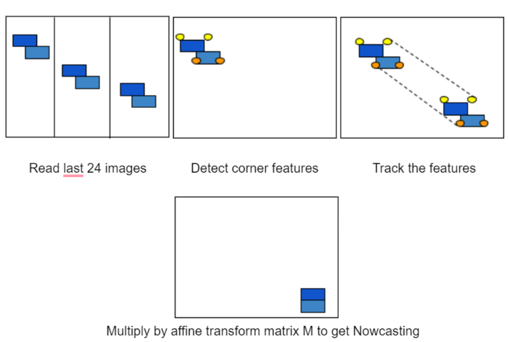


Fig. 3. Precipitation Nowcasting technique

VIII. Corner Detection Algorithm

Corner are points in image where slight shift in location will result in a large change in intensity in both the axes.Let I be the image intensity and f be the sum of squared differences between two patches.Let the window under consideration be W.

$$f(\Delta x, \Delta y) = \sum (I(x_k, y_k) - I(x_k + \Delta x, y_k + \Delta y))^2$$

Using the Taylor expansion and linearity we get

$$I(x + \Delta x, y + \Delta y) \approx I(x, y) + I_x(x, y)\Delta x + I_y(x, y)\Delta y$$

$$f(\Delta x, \Delta y) \approx \sum (I_x(x, y)\Delta x + I_y(x, y)\Delta y)^2$$

$$f(\Delta x, \Delta y) \approx (\Delta x, \Delta y)M(\Delta x, \Delta y)^T$$

$$where\ M = \sum \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

By solving for the eigenvectors and values of M we can calculate the R-score. If the $\lambda_1 >> \lambda_2$ we get an edge as there is an increase in SSD(sum of squared differences) along one direction. If eigenvalues are large it's a corner as there is a sharp increase in SSD in almost all the directions and a flat region if eigenvalues are small.
R score(Harris corner detector)=$\lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2$
R score(Shi Tomasi corner detector)=$min(\lambda_1, \lambda_2)$
In practice it is seen that shi Tomasi performs better than harris corner detector.

IX. Errors in Lucas Kanade & Future Research

There are some cases where the algorithm will fail badly i.e when our assumptions of Brightness constancy are not satisfied, the motion that we are trying to detect is not small and Neighbouring points don't move in the same way(window size is too large).Some practical cases include a rotating sphere, a barber pole that has alternate coloured stripes and changing light intensities that can make things seem to move due to changing brightness. Still, Lucas Kanade has tremendous application in practical uses like improving video quality, image segmentation, tracking objects, recognizing events and activities.
Traditional optical flow algorithms are iterative algorithms, have convergence issues and are difficult to optimize further in terms of accuracy. Currently Deep Learning techniques like FlowNet(3D CNN) and supervised and unsupervised learning algorithms. Recently FlowNet2 was introduced which had 50% less estimation error compared to FlowNet which stacks several neural nets and uses warping techniques to predict displacement.It was a very fascinating topic to study which uses simple Taylor series expansion to yield such a beautiful algorithm and has applications in so many fields. [1]–[6]

References

[1] Takeo Kanade Bruce D. Lucas. An iterative image registration technique with an application to stereo vision. *,Proc of 7th Joint Conf on Artificial Intelligence(IJCAI)*, pages 674–679, 1981.
[2] Maik Heistermann Georgy Ayzel and Tanja Winterrath. Optical flow models as an open benchmark for radar-based precipitation nowcasting (rainymotion v0.1). 9 April,2019.
[3] David J. Fleet and Y. Weiss. *Optical Flow Estimation*. 2006.
[4] David J. Fleet and Y. Weiss. https://www.cs.toronto.edu/fleet/research/Papers/flowChapter05.pdf, 2005. Accessed 31-March-2022.
[5] Instructor: Dr. Mubarak Shah UCF CRCV. Lecture 10 - Lucas- Kanade Tracker (KLT)(Sep 30, 2012). https://www.youtube.com/watch?v=tzO245uWQxA. [Online; accessed 31-March-2022].
[6] Ranjay Krishna Juan Carlos Niebles. Lecture:motion. https://vision.stanford.edu/teaching/cs131_fall1718/files/17_motion.pdf. Accessed 31-March-2022.