INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY

HYDERABAD

Assignment 2 Solutions

Course: Computer Vision
Name: K.N.Amruth sagar
Roll.NO: 2021701013

1 Basics of Optical Flow

1.1 Thought experiments

Question 1
The number-one use of optical flow in visual effects is for retiming a se- quence — speeding it up or slowing it down. Describe how optical flow could be used to create slow motion video. You can find the answer in Amazing Slow Motion Videos With Optical Flow video on YouTube. It's movie time now! Let's watch two epic movie clips from two of the Academy Award winning movies for the Best Visual Effects - What Dreams May Come (1998), and The Matrix (1999).

Solution

Slow motion videos are generally captured at very high frame rates, some cameras like Phantom v2512 can reach upto *one million* frames per second. In order to make a slow motion video from a source video whose frame rate is not high, we need to interpolate and find the intermediate frames which will increase the framerate of the video, and if the interpolated video is played at the same initial frame rate, it will look like a slow motion video as there are more number of frames.

There are many ways of filling in the intermediate frames in a video, and one common way other that the optical flow, which is our interest here, is the frame blending. It just takes the average of the two neighbouring frames, and then cross-dissolves to smooth out the artifacts in the video. If there is not much motion, this is a reasonable method, but it will not provide satisfactory results.

Optical flow: Using the motion constraint equation in optical flow, we can estimate the motion of each pixel from consecutive frames and can generate an intermediate image using the motion information. Using the motion of each pixel we can warp the previous frame such that the new frame is the intermediate frame of the two consecutive frames.This has to be done in a dense manner, as the optical flow method cannot know what is the foreground and what is background.

Is it always good to use optical flow?: If the subjects in the video are moving slowly, then there will be almost no artifacts in the interpolated video, but if the video contains sudden movements or any quick cross-frame movements, we can easily spot the ghosting artifacts in the video. An another case is when different things in the frame have similar contrast and color, which will introduce warping effects between them if we apply optical flow.

Question 2
In The Matrix, one of the most remembered, iconic use of the tactic comes during the rooftop scene where Neo effortlessly dodges one bullet after an- other. (Re-)watch "Bullet Time" here and explain briefly how optical flow is used. You may also find this amazing video on bullet-time interesting.

Solution

According to this *Bullet time* reference, they used a mixture of techniques, but the core technique to produce the awesome effect is *Optical flow*. Before shooting the actual scene, tests were conducted in which 27 cameras were used and the positions for the cameras and the animation for the slow motion bullets were tested.

Just using optical flow information between the frames and intertiming them, and using it to make the slowmotion effect was really hard and didn't result in a very good visual effect, and was particularly difficult to handle the main actor even when the action was filmed with a green screen. To achieve the effect properly, more cameras, optical flow, frame interpolation and other techniques were employed later on, which resulted in the oscar winning effect.

The effect was achieved by decomposing the actor into limbs, using optical flow to do interpolation based on the motion data of consecutive frames, and generating the intermediate frames and stabilizing the parts, and then rebuilding the actor again by combining the stabilized parts of the actor. In matrix "Bullet time" scene, ghosting artifacts are visible sublty near the left hand part. Nevertheless, this technique is a huge achivement considering both the time and computing resources available at that time.

> **Question 3**
> So breathtaking, heartbreaking and brimming with emotion, WDMC is a journey into the afterlife and deals with a dead man's attempt to reunite with his wife. Catch a glimpse of the ”Painted World” here! You'll now describe briefly on how optical flow is used to create this “painterly effect”.

Solution

Once again, optical flow technique was used, not to do interpolation, but to generate painting-like afterlife frames from the actual shot. This elaborate but genius process involves calculating the optical flow between the frames and obtaining the information in terms of x and y shift for each and every pixel, then using a partical system that takes the optical flow as transformation information, along with other information like orientation, depth and other rules to draw brush strokes of colors to produce painting like output.

The output can be seen as the combination of various maps superimposed on each other with a particular alpha value, depending on the importance given to that particular transformation. The particle system and the optical flow information were the key components, because one is the whole pipeline that worked on various data and rules and then the other being the method that provides with info that changes from frame to frame.

Reference: WDMC painted world

> **Question 4**
> Consider a Lambertian ball that is: (i) rotating about its axis in 3D under constant illumination and (ii) stationary and a moving light source. What does the 2D motion field and optical flow look like in both cases.

Solution

CASE 1: *Rotating lambertian ball in constant illumination*
As the ball is rotating about its axis, there will be no movement of pixels i.e. the surface of the ball moves while the center remains still and we will not observe any optical flow because there is no intensity change between the frames and as the illumination is constant, and the surface of the ball is lambertian, which means it is perfectly matte and will have a constant reflectance. Now the equation to calculate u and v becomes,

$$\begin{bmatrix} I_x(p_1) & I_y(p_1) \\ I_x(p_2) & I_y(p_2) \\ \vdots & \vdots \\ I_x(p_n) & I_y(p_n) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -\cancelto{0}{\begin{bmatrix} I_{t1} \\ I_{t2} \\ \vdots \\ I_{tn} \end{bmatrix}}$$

The only way the LHS side of the equation equates to 0 is when the optical flow is 0 because the image gradients cannot be zero at the edges of the sphere.

CASE 2: *Still lambertian ball in varying illumination*
In varying illumination, even though the ball is still, the intensity of the pixels between the frames changes which leads to non-zero optical flow, because the gradients vary and are not completely 0 and there is a difference in intensity between the frames, which will give a non zero solution for the displacements u and v.

Let us assume a case where the light source is rotating around the ball with the center of rotation on the axis of the sphere, then if we visualize the optical flow, the motion vectors move along the changing intensities.

1.2 Concept Review

Question 1
List down the important assumptions made in optical flow estimation. Describe each one of them in one-two lines.

Solution

The important assumptions made in optical flow estimation are:

Assumption 1: *Brightness constancy assumption*, which assumes that the intensity of a points will remain constant for a small change in position over time. It can be written as

$$I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t)$$

Assumption 2: The change in the position of the point and the time step are very small. This assumption lets us expand the change in intensity as a taylor series,

$$f(x + \delta x) = f(x) + \frac{\partial f}{\partial x}\delta x + \frac{1}{2!}\frac{\partial^2 f}{\partial x^2}\delta x^2 + \cdots$$

similarly, in multivariate case,

$$\text{let } [\delta x, \delta y, \delta t] = X_0$$

$$I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t) + \nabla I(x, y, t)X_0 + \frac{1}{2!}X_0^T \nabla^2 I(x, y, y)X_0 + \cdots$$

$$\cancel{I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t)} + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t$$

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

Question 2
Formalize the objective function of the classical optical flow problem. Clearly mark the data term and the spatial term. What does the ob- jective function imply about the noise distribution?

Solution

The objective function of the classical optical flow problem is to estimate the optical flow components u, v which minimize the error between the estimate and the actual values. Considering the image as a function of position and time, $I(x, y, t)$, then according to the brightness constancy assumption, the intensity of the image should not change for tiny steps in position and time,

$$E = \sum \left(I(x + \delta x, y + \delta y, t + \delta t) - I(x, y, t)\right)^2$$

using taylor series approximation,

$$E \approx \sum \left(\cancel{I(x, y, t)} + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t - \cancel{I(x, y, t)}\right)^2$$

Question 3
In optimization, why is the first-order Taylor series approximation done?

Solution

Let the change in intensity of a pixel through time be $I(x, y, t)$. This function takes in 3 variables, and tells us the intensity of the pixel at the given coordinate positions (x, y) and at time t. Now this function can be very complex and we have to know the equation in order to know the intensity of the pixels after a little displacement and next time frame.

This is where the *first order Taylor series approximation* helps. Taylor series goes on until the function becomes constant and the remaining derivatives become zero. If we take the first order taylor aproximation of a function $f(x)$ then it becomes,

$$f(x + \delta x) = f(x) + \frac{\partial f}{\partial x}\delta x + \frac{1}{2!}\frac{\partial^2 f}{\partial x^2}\delta x^2 + \cdots$$

taking the first order approximation of the above function

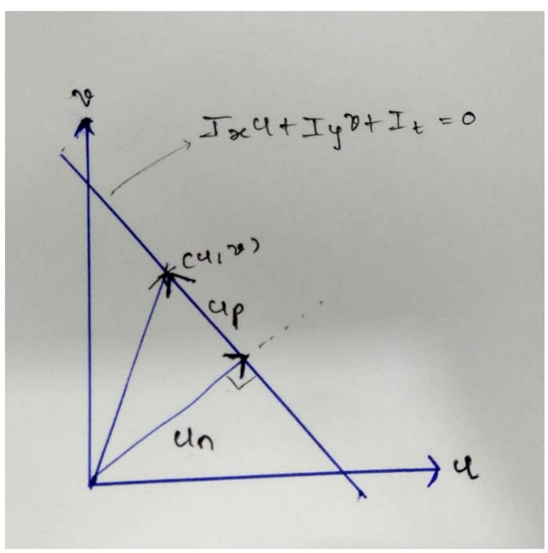
$$f(x + \delta x) \approx f(x) + \frac{\partial f}{\partial x}\delta x$$

The approximation holds very well when the change in x is very small, because, *the first approximation and the original function coincide as the change in x (δx) becomes smaller and smaller.* This property is what we want when we assumed that the brightness of the pixel will not change for very small displacements in x and y direction and over a small time step δt. The left side of the equation becomes the same as a term in right side of the linear approximation,

$$\cancel{I(x + \delta x, y + \delta y, t + \delta t)} = \cancel{I(x, y, t)} + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t$$

Question 4
Geometrically show how the optical flow constraint equation is ill-posed. Also, draw the normal flow clearly.

Solution



In the above diagram, the axes are u and v. The true flow values are $\mathcal{U} = (u, v)$. The line passing through the flow values is $I_x u + I_y v + I_t = 0$, which is the optical flow constrained equation. The equation is ill-posed because, if we consider a particular pixel and then calculate the above equation, we will not get the actual flow information because there are *infinitely many values of u and v which satisfy the equation.*

Moreover, *we can only get the* ***normal flow*** *locally.* The true optical flow $\mathcal{U}$ can be decomposed into its normal vector $\mathcal{U}_n$ and a parallel component to a line $\mathcal{U}_p$. The normal component can be easily calculated as $\frac{|I_t|}{\sqrt{I_x^2 + I_y^2}}$, but we cannot calculate the parallel component $\mathcal{U}_p$, as it can be any point on the given line.

So, as the constrained equation is underdetermined system, means there are less equations than the unknowns, we can only get the normal flow locally and this is also the reason behind the aperture problem.

1.3 Analyzing Lucas-Kanade Method

Question 1
Why is optical flow only valid in the regions where local structure-tensor $A^T A$ has a rank 2? What role does the threshold τ play here?

Solution

The matrix A consists of x and y gradients of the points in the considered neighbourhood.

$$A = \begin{bmatrix} I_x(p_1) & I_y(p_1) \\ I_x(p_2) & I_y(p_2) \\ \vdots & \vdots \\ I_x(p_{n^2}) & I_y(p_{n^2}) \end{bmatrix}_{n^2 \times 2}$$

Now the product $A^T A$ becomes

$$A^T A = \begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix}$$

We can only calculate the optical flow where the local-structure tensor has rank two, because, when we say $A^T A$ has a full rank of 2, *then it means it is a symmetric positive-definite matrix*, which means that the eigen values are both positive and non-zero, and as the determinant is the product of eigen values, the determinant is also > 0.

By having a full rank for the local-structure matrix, we can calcualte the solution as the matrix is invertible as the determinant is non zero. The optical flow is calculated as follows when the rank is full,

$$[u, v]^T = -(A^T A)^{-1} A^T b$$

The role of the treshold τ here is to check whether the matrix $A^T A$ has rank 2 or not. If the smallest eigen value is smaller than the treshold $\tau = 0.01$, then there is no need of computing optical value in that region.

Question 2
In the experiments, did you use any modifications and/or thresholds? Does this algorithm work well for these test images? If not, why?

Solution

In experiments, if we decrease the value of the treshold τ lesser than 0.01, for example at $\tau = 0.005$, then we end up calculating optical flow at more regions than before, the algorithms works well but there is no reason for selecting the points with eigen values very small, which will add nothing to the optical flow calculation as we use it to check the rank of the local-structure tensor and not calculate optical flow at those regions which might be regions which are plain or contain edges.

if we increase the treshold, for example at $\tau = 0.05$, only few points are selected, but the optical flow calculated at those regions give very good response, because we are tresholding and rejecting all the points whose smallest eigen value is less than that. These points will almost be corners because the corners have bigger eigen values, and we get rid of points which are more edge-like at greater thresholds.

Question 3
Try experimenting with different window sizes. What are the trade-offs associated with using a small versus a large window size? Can you explain what's happening?

Solution

Taking a smaller window:

When we only take a small window, we will consider a very small neighbourhood, which is *good when the velocity of the center pixel/corner is very small.* Then we get better estimates for the slow moving points as we are not considering many points that might add noise

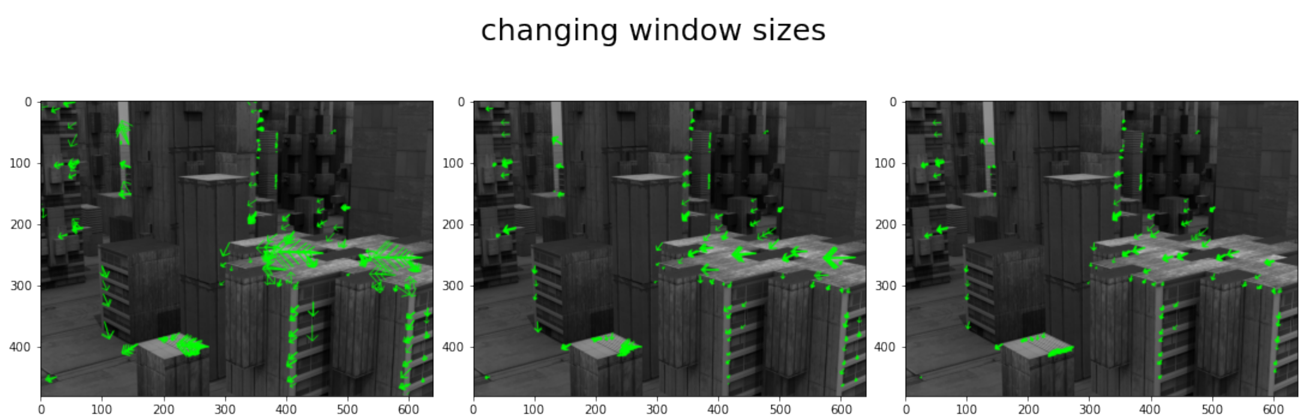
The problem arises when the corner moves so fast, that the center pixel is completely outside of the window in the next frame, which will give us completely wrong esitmates of optical flow which when visualized give us completely random directions for the optical flow.

Taking a bigger window:

When we take a larger window, we will consider a larger neighbourhood which is *good for the corner/interest points moving very fast, which result in huge translation from frame to frame.* Larger windows gives us better estimates for fast moving interest points.

The problem is the larger window might also add so much noise, because there can be cases where the large window is also considering other interest points, which might be moving in various directions, which can give erroneous estimates for slow moving points or regions which include objects moving in opposite directions.

The below image shows the outputs for the window sizes of $(5, 5)$, $(25, 25)$ and $(50, 50)$



> **Question 4**
> There are locations in an image where Lucas–Kanade optical flow will fail, regardless of choice of window size and sigma for Gaussian smoothing. Describe two such situations.You are not required to demonstrate them.

Solution

There are two situations in which no window size or sigma is beneficial.

Case 1: Regions with uniform intensity: Regions with uniform intensity will result in a lower rank local-structure tensor $A^T A$ which is due to very small values (almost 0) in these regions as the change in the intensities are almost non-existant. In these cases the choice of window size or sigma doesn't matter.

Case 2: Regions with edges: Choice of window size and sigma also doesn't matter for regions where there is an edge, as one of the eigen value will be very close to 0 (will be 0 in an ideal case, which lowers the rank of matrix), meaning that the optical flow cannot be estimated properly for such cases.

Question 5
Did you observe that the ground truth visualizations are in HSV colour space? Try to reason it.

Solution

The visualizations of the optical flow are done in HSV instead of RGB because,

- *We can represent the angle of the optical flow with the hue value in HSV*, which denotes different color at differeWnt angles, starting from Red at 0°, Yellow at 60°, Green at 120°, Cyan at 180°, Blue at 240°, Magenta at 300° and again ending at Red.

- *We can represent the magnitude of the optical flow vector with saturation*, low saturation means slow movement and high saturation (pure color) means the point's movement is fast.

The main reason for using HSV other than using Hue and saturation components is that we can comprehend colors in HSV space like real life where we use intensity in our color description, which is the V(value) component and the color is the hue component.

Using RGB space to visualize optical flow vectors is a very painful task, as we have to set ranges for different values for u and v, their angle and their magnitude. RGB is just a triplet of numbers which says nothing about the brightness/intensity of the color, or atleast tell what is the pure color.

In HSV, for example, if the color at a particular pixel is light blue, we can understand that the optical flow direction is in the bottom-left(240°) direction, and the lightness says that the saturation is low, which means the velocity is also low. It is impossible to know the optical flow if it were visualized in RGB.