CV Assignment - 2 : The Secrets of Optical Flow

Name: T. Gokul Vamsi

Roll no: 2019111009

Branch: CSD

1. Basics of Optical Flow

1.1 Thought Experiments

1) Optical flow is a very useful technique that could be used to create slow motion video. Essentially, a slow motion video is supposed to contain duplicated frames from the original video, so as to achieve the desired slow motion effect, increasing the video's duration.

Using optical flow, consecutive frames of a video can be considered, to obtain motion vectors for pixels of these frames; and multiple duplicate frames can be interpolated between them, due to the motion understanding obtained. Thus, such new (interpolated) frames constructed by this optical flow algorithm, are integrated seamlessly into the original video, causing the slow motion effect.

2) In the iconic scene of 'The Matrix (1999)', where Neo dodges bullets, the camera seems to move quickly around Neo to capture the entire scene in a smooth and intriguing manner. Optical flow is one of the key concepts involved in this creation.

Several cameras are placed such that they surround Neo, and are triggered appropriately to capture the frames at short intervals, of the entire circular (360°) view. These captured frames, are then interpolated to incorporate multiple frames, between ~~within~~ such consecutive frames, to make the motion appear more smooth and less jittery. This interpolation is achieved by optical flow, as the motion understanding it obtains are utilized to create such an effect (smooth slow-motion-transitions).

---

3) In the given scene of WDMC, where a heaven is modelled, there is a 'painterly effect' to various objects in the scene, and optical flow is a central aspect utilized, to achieve this effect.

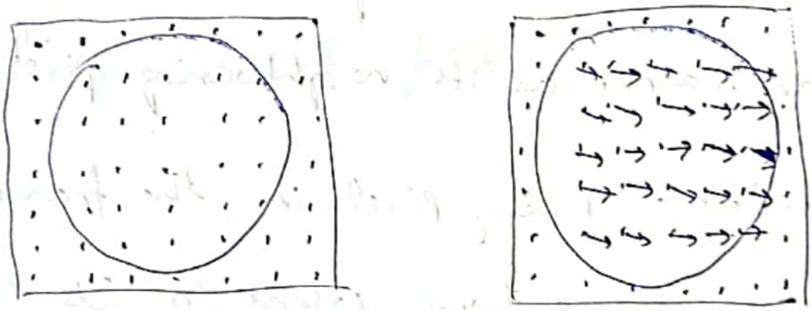
The normal scene (with flowers, trees, etc.) is considered, and paint brushstrokes are incorporated into it digitally, to achieve frames in desired format such that the objects in the scene actually have this paint

The motion of these pixels with point is tracked via the motion vectors computed by the optical flow algorithm, thereby creating an effect of the scene having movements due to the motion interpolation; creating the 'pointerly' effect.

---

4)

(i) For a Lambertian ball rotating about its axis in 3D under constant illumination, the optical flow would look completely stationary, and no motion vectors are observed as the light source is stationary. Although the 2D motion field would clearly indicate motion, as the ball is rotating. They can be roughly depicted by below diagrams:



Optical flow (stationary) 2D motion field (movement)

(ii) For a Lambertian ball that is stationary and the light source is moving, the action would yield reversed results as in previous case. The optical flow would depict motion in the pixels, as it comprehends the motion of the light source as the motion of pixels corresponding to the ball, due to change in illumination. Although, the

2D motion field would not indicate any such motion, as the ball is stationary in the scene. Therefore, both the images are reversed in this scenario relative to the previous case, as the static image would correspond to 2D motion field, and the image with motion depicts optical flow. (Refer previously drawn images - reversed action).

1.2 Concept Review

1) Important assumptions made in optical flow estimation are:

→ Brightness constancy: The brightness of pixels (intensity) remains constant in the projected frame.

→ Spatial coherence: Every individual pixel moves in a similar manner as its neighbouring pixels.

→ Small motion: Every pixel in the frame is assumed not to move by a large extent in its corresponding projected frame.

2) The objective function of the classical optical flow problem is given by:

$$E(u,v) = \sum \Bigg(\overbrace{\rho_D\big(I_1(i,j) - I_2(i+u_{ij},\, j+v_{ij})\big)}^{\boxed{\text{Data term}}} +$$

$$\underbrace{\lambda\Big[\rho_s(u_{ij}-u_{i+1,j}) + \rho_s(u_{ij}-u_{i,j+1}) + \rho_s(v_{ij}-v_{i,j+1}) + \rho_s(v_{ij}-v_{ij+1})\Big]}_{\boxed{\text{Spatial term}}} \Bigg)$$

In this equation,

$u \rightarrow$ horizontal component of optical flow field

$v \rightarrow$ vertical component of optical flow field

$I_1, I_2 \rightarrow$ Images

$u_{ij}, v_{ij} \rightarrow$ Elements of u, v (i^{th} row, j^{th} column)

$\lambda \rightarrow$ Regularization parameter

$\rho_D \rightarrow$ Data penalty function

$\rho_S \rightarrow$ Spatial penalty function

As marked in the equation clearly, the data term corresponds to the difference in intensities, as it contains brightness information / data; and the spatial term corresponds to spatial coherence, where pixel motion is compared with its neighbours; and necessary difference values are computed. Penalty functions are used for both the terms appropriately.

Based on the nature of the objective function which includes linear terms along with some penalty functions, the noise distribution can be reppresented by a multi-dimensional Laplace distribution (as it involves Taylor expansion, difference in intensities and motion vectors).

Scanned with CamScanner

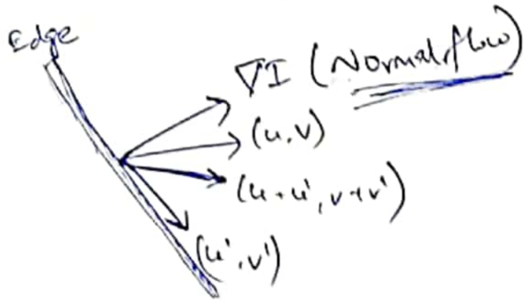
3) In optimization, first-order Taylor series approximation is done because of the assumption that the motion of pixels in considered frame is small; and hence the approximation would fairly complement this assumption. It further mathem[illegible] simplifies (improves) the efficiency of computation, as only linear te[illegible] are considered in the constraint indicating motion between two frames is small.

---

4) It can be observed that the optical flow constraint is ill-posed because of the ambiguity associated with motion estimation. Let the pixel intensities at time steps t and $(t+1)$ be given by I_t. Considering motion vectors $[u\ v]$, we know the constraint equation to be:

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

Clearly, we can observe that, the component of motion vector of pixel under consideration, which is perpendicular to the gradient, cannot be accurately calculated. Consider the following diagram which represents this geometrically:



Here, as the component parallel to the edge is not relevant for any motion vector $[u,\ v]$ that satisfies the optical flow constraint equation, the corresponding vector $[u+u' \quad v+v']$ also satisfies the equation. where $[u' \quad v']$ is a vector along the direction of the edge, i.e. perpendicular to the gradient, due to the fact that $\nabla I \cdot [u' \quad v']^T = 0$

Thus, these numerous geometrical solutions explain the ill-posed nature of the constraint equation.

2. Single-Scale Lucas-Kanade Optical Flow:

2.3 Analyzing Lucas-Kanade method

1) For obtaining the least squares solution of Lucas-Kanade equation, we solve the following:

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

In cases where local structure-tensor $A^T A$ has rank = 2, it is clearly able to assist in obtaining a valid solution $\begin{bmatrix} u \\ v \end{bmatrix}$, as $A^T A$ is invertible, and $(A^T A)^{-1} A^T b$ would generate required motion vector. But if rank $(A^T A) < 2$ (it clearly can't be greater than 2 due to its dim) $A^T A$ is no longer invertible, as $|A^T A| = 0$.

Thus, optical flow is valid only in regions where local structure-tensor $A^T A$ has rank = 2.

Considering threshold 'T', it defines if a motion vector $\begin{bmatrix} u \\ v \end{bmatrix}$ can be computed for a particular pixel. ~~If~~ rank $(A^T A) < 2$, it clearly indicates that the lower eigen value would be 0, which would definitely be less than any positive threshold T. Even in cases where the smaller eigen value is lesser than T, it is considered to be affected by noise, and could be considered that the motion vector cannot/shouldn't be computed for that pixel. Thus, the threshold 'T' plays a key role in discarding pixels where optical flow won't be valid.

---

2) In the experiments, the following variations/thresholds were incorporated.

→ If smallest eigen value of A^TA (local-structure tensor) is less than a threshold τ, its optical flow isn't computed

→ A weighted Gaussian kernel was used to penalize pixels away from centre of the window (weighted-version of least squares solution).

It was observed that error for least squares solution was ~~greater~~ Lesser than that obtained on using weighted version (due to irregular motion). The algorithm seemed to work well enough for the given image (building of urban area), although the error was still on the higher side, due to the fact that there are more number of corners in the image which move in a relatively asymmetric manner, i.e, motion is small for few regions and seems significant for few other regions. This indicates how spatial coherence and brightness constancy constraints are affected, reducing the accuracy of Lukas-Kanade on these test images.

---

3) On experimenting with different window sizes, it was observed that there is a tradeoff associated with it. If we consider a smaller window size, it is more prone to noise, as we are considering only a very small portion around a pixel, to determine its motion vector. Thus, the least squares solution corresponds to a smaller number of constraints, indicating that the motion vector calculated would not be as accurate.

If we consider a large window instead, it alleviates the problem of noise significantly, but as it stems from the assumption that all pixels in the particular window have the same motion vector, it leads to an over-constrained system of equation, where clusters of pixels are found to have the same motion vector. This fails in situations where a point/pixel doesn't move like its neighbours.

Thus, it is essential to achieve a balance between such window sizes, to obtain good results.

---

4) Two situations where Lucas-Kanade optical flow will fail irrespective of window size and sigma, are explained as follows:

→ when there are particular regions in the image where the motion is large, Lucas-Kanade optical flow will fail, as the Taylor approximation won't be valid for this region. For instance, if we consider two frames of a very fast-moving train, optical flow wouldn't be effective in motion vector computation irrespective of window size and sigma, as Lucas-Kanade algorithm considers only a small local space around the pixel; and cannot capture motion effectively.

→ In cases where there is a significant change in the illumination, Lucas-Kanade approach would fail, as the brightness constancy assumption is not satisfied by these regions. For such regions where the illumination of pixels in first frame is very different from the illumination intensity of corresponding pixels in the next time step (for a fixed window and sigma), the difference in intensities would be significant (between $I(x,y,t)$ and $I(x+u, y+v, \ldots)$ which affects the accuracy in these regions.

---

5) It was observed that the ground truth visualizations are in HSV colour space, as it assists in the partioning between the image intensity and colour of the image. This is done so as to improve the efficiency/performance

as well as to eliminate any noise due to the chromatic information corresponding to the image. As only the intensity is of primary relevance, it is segregated from the colour information, to yield better quality output more efficiently.