HOMEWORK 2: LUCAS-KANADE TRACKING

16-720A Computer Vision (Spring 2023)
https://canvas.cmu.edu/courses/32966

OUT: Feb 13th, 2023
DUE: March 3rd, 2023 11:59 PM
Instructor: Deva Ramanan
TAs: Kangle Deng, Vidhi Jain, Xiaofeng Guo, Chung Hee Kim, Ingrid Navarro

Instructions

- Please refer to the course logistics page for information on the **Collaboration Policy** and **Late Submission Policy**.
- **Submitting your work:** There will be two submission slots for this homework on **Gradescope**: Written and Programming.
  - **Write-up.** For written problems such as short answer, multiple choice, derivations, proofs, or plots, we will be using the written submission slot. Please use this provided template. We don't accept handwritten submissions. Each answer should be completed in the boxes provided below the question. You are allowed to adjust the size of these boxes, but make sure to link your answer to each question when submitting to Gradescope. Otherwise, your submission will not be graded.
  - **Code.** You are also required to upload the code that you wrote to solve this homework, to the Programming submission slot. Your code may be run by TAs so please make sure it is in a working state. Also, make sure to use appropriate names for your variables and to add comments and explanations to your code for the TAs to understand it better. The assignment must be completed using Python 3 in Jupyter. We recommend setting up a conda environment, but you are free to set up your environment however you like.
  - Regrade requests can be made after the homework grades are released, however this gives the TA the opportunity to regrade your entire paper, meaning if additional mistakes are found then points will be deducted.
- **Start early!** This homework may take a long time to complete.
- **Verify your implementation as you proceed.** If you don't verify that your implementation is correct on toy examples, you will risk having a huge mess when you put everything together.
- **Q&A.** If you have any questions or need clarifications, please post in Slack or visit the TAs during office hours. Additionally, we provide a **FAQ** (section 6) with questions from previous semesters. Make sure you read it prior to starting your implementations.

Overview

Lucas-Kanade (LK) is a widely known vision-based method for tracking features in image sequences. In a nutshell, it uses the notion of optical flow for estimating the motion of pixels in subsequent images. This homework explores the core aspects for implementing the LK tracking method and efficiency improvements.

What you'll be doing:

This homework is worth 100 points + 20 extra credit points, and consists of four parts:

1. In section 1, you will implement a Lucas-Kanade (LK) tracker for **pure translation with a single template**. You will explore two methods for performing template updates: a naive update, and an update with drift correction. You will test your implementation on two sequences `carseq.npy` and `girlseq.npy` which we provide in the `data` folder. This section is worth 50 points.
2. In section 2, you will modify the tracker to account for **affine motion**, and you will then implement a motion subtraction method to track moving pixels on a scene. You will test your implementation on two sequences `antseq.npy` and `aerialseq.npy` which we provide in the `data` folder. This section is worth 35 points.
3. In section 3, you will implement **efficient** tracking via inverse composition. You will also test your implementation `antseq.npy` and `aerialseq.npy`. This section is worth 15 points.
4. For **extra credit**, in section 4 you're asked to run your implementation on a video of your choice. This section is worth 20 extra points.

Resources:

In addition to the course materials, you may find the following references useful;

- Simon Baker, et al. *Lucas-Kanade 20 Years On: A Unifying Framework: Part 1*. CMU-RI-TR-02-16, Robotics Institute, Carnegie Mellon University, 2002. [link]
- Simon Baker, et al. *Lucas-Kanade 20 Years On: A Unifying Framework: Part 2*. CMU-RI-TR-03-35, Robotics Institute, Carnegie Mellon University, 2003. [link]

1 Lucas-Kanade Tracking (50 total points)

In this section, you will implement a simple Lucas-Kanade (LK) tracker with a single template.

Coding questions for **part 1** must be implemented in `LucasKanade.ipynb`. We provide you with starter code for each question, as well as default parameters. To test your tracker, you will use `carseq.npy` (top row in fig. 1.1) and `girlseq.npy` (bottom row in fig. 1.1). You can find these files in the `data` folder.

Problem Formulation. Following the notation in [1], let us consider a tracking problem for a 2D scenario. We refer to $\boldsymbol{I}_{1:T}$ as a sequence of T frames, where $\boldsymbol{I}_t$ is the current frame and $\boldsymbol{I}_{t+1}$ is the subsequent one. We represent a **pure translation** warp function as,

$$\mathbf{x}' = W(\mathbf{x}; \mathbf{p}) = \mathbf{x} + \mathbf{p} \tag{1.1}$$

where $\mathbf{x} = [x, y]^T$ is a pixel coordinate and $\mathbf{p} = [p_x, p_y]^T$ is an offset.

Given a template, T_t in frame, $\boldsymbol{I}_t$ which contains D pixels, the Lucas-Kanade tracker aims to find an offset $\mathbf{p}$ by which to translate the template on $\boldsymbol{I}_{t+1}$, such that the squared difference between the pixels on those two regions is minimized,

$$\mathbf{p}^* = \underset{\mathbf{p}}{\operatorname{argmin}} \sum_{\mathbf{x}\in T_t} ||I_{t+1}(W(\mathbf{x}; \mathbf{p})) - T_t(\mathbf{x})||^2 = \underset{\mathbf{p}}{\operatorname{argmin}} \left\| \begin{bmatrix} I_{t+1}(W(\mathbf{x}_1; \mathbf{p})) - T_t(\mathbf{x}_1) \\ \vdots \\ I_{t+1}(W(\mathbf{x}_D; \mathbf{p})) - T_t(\mathbf{x}_D) \end{bmatrix} \right\|^2 \tag{1.2}$$

In other words, we are trying to align two patches on subsequent frames by minimizing the difference between the two.

1.1 Theory Questions (5 points)

Starting with an initial guess for the offset, *e.g.* $\mathbf{p} = [0, 0]^T$, we can compute the optimal $\mathbf{p}^*$, iteratively. In each iteration, the objective function is locally linearized by the first-order Taylor expansion,

$$\boldsymbol{I}_{t+1}(\mathbf{x}' + \Delta\mathbf{p}) \approx \boldsymbol{I}_{t+1}(\mathbf{x}') + \frac{\partial \boldsymbol{I}_{t+1}(\mathbf{x}')}{\partial \mathbf{x}'^T} \frac{\partial W(\mathbf{x}; \mathbf{p})}{\partial \mathbf{p}^T} \Delta\mathbf{p} \tag{1.3}$$

where $\Delta\mathbf{p} = [\delta p_x, \delta p_y]^T$, is the delta change of the offset, and $\frac{\partial \boldsymbol{I}_{t+1}(\mathbf{x}')}{\partial \mathbf{x}'^T}$ is a vector of the x and y image gradients at pixel coordinate $\mathbf{x}'$. We can then take this linearization into equation 1.2 into a vectorized form,

$$\underset{\Delta\mathbf{p}}{\arg\min} ||\mathbf{A}\Delta\mathbf{p} - \mathbf{b}||^2 \tag{1.4}$$

such that $\mathbf{p} \leftarrow \mathbf{p} + \Delta\mathbf{p}$ at each iteration.

The questions that follow will be useful for your implementation. Please answer each of them in their corresponding box. The answers should be short.

Q1.1.1 What is $\frac{\partial W(\mathbf{x};\mathbf{p})}{\partial \mathbf{p}^T}$? (**Hint:** It should be a 2x2 matrix)

Answer for Q1.1.1

For a pure translational warp matrix,
$\frac{\partial W(x;p)}{\partial p^T} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Q1.1.2 What is **A** and **b**?

Answer for Q1.1.2

The squared difference

$p^* = argmin_p \sum \left\| I_{t+1}\left(W(x;p)\right) - T_t(x) \right\|_2^2$

Given $I_{t+1}\left(W(x;p)\right) = I_{t+1}(x') + \frac{\partial I_{t+1}(x')}{\partial x'^T}\frac{\partial W(x;p)}{\partial p^T}\Delta p$

$p^* = argmin_p \sum \left\| I_{t+1}(x') + \frac{\partial I_{t+1}(x')}{\partial x'^T}\frac{\partial W(x;p)}{\partial p^T}\Delta p - T_t(x) \right\|_2^2$

$p^* = argmin_p \sum \left\| \frac{\partial I_{t+1}(x')}{\partial x'^T}\frac{\partial W(x;p)}{\partial p^T}\Delta p - (T_t(x) - I_{t+1}(x')) \right\|_2^2$

Comparing with $argmin_p \sum \|A\Delta p - b\|_2^2$, $A = \frac{\partial I_{t+1}(x')}{\partial x'^T}\frac{\partial W(x;p)}{\partial p^T}$, $b = T_t(x) - I_{t+1}(x'))$

Q1.1.3 What conditions must $\mathbf{A}^T\mathbf{A}$ meet so that a unique solution to $\Delta\mathbf{p}$ can be found?

Answer for Q1.1.2

To minimize Δp, $\frac{\partial p^*}{\partial \Delta p} = 0$

$\sum 2A^T(A\Delta p - b) = 0$

$\Delta p = \sum (A^T A)^{-1} A^T b$

So, for Δp to obtain a unique solution, $A^T A$ must be invertible and non-singular matrix

1.2 Lucas-Kanade (15 points)

Implement the function,

```
p = LucasKanade(It, It1, rect, threshold, num_iters, p0 = np.zeros(2))
```

where, `It` is the image frame, $\boldsymbol{I}_t$; `It1` is the image frame $\boldsymbol{I}_{t+1}$; `rect` is a 4-by-1 vector defined as $[x_1, y_1, x_2, y_2]^T$ that represents the corners of the rectangle comprising the template in $\boldsymbol{I}_t$. Here, $[x_1, y_1]^T$ is the top-left corner and $[x_2, y_2]^T$ is the bottom-right corner. The rectangle is inclusive, i.e., it includes all four corners; `p0` is the initial parameter guess $[\delta p_x, \delta p_y]^T$. Your optimization will be run for `num_iters`, or until $||\Delta\mathbf{p}||_2^2$ is below a `threshold`.

Your code must **iteratively** compute the optimal local motion ($\mathbf{p}^*$) from frame I_t to frame I_{t+1} that minimizes eq. (1.2). As you learned during lecture, at a high-level, this is done by following these steps:

1. Warp the template;
2. Build your linear system (**Q1.1.2**);
3. Run least-squares optimization (eq. (1.4));
4. Update the local motion.

Note: For this part, you will have to deal with fractional movement of the template by doing interpolations. To do so, you may find Scipy's function `RectBivariateSpline` useful. Read the documentation for `RectBivariateSpline`, as well as, for evaluating the spline `RectBivariateSpline.ev`.

Though we recommend using the aforesaid function, other similar functions for performing interpolations can be used as well. See the FAQ (section 6) for more details.

Include the code you wrote for this part in the box below:

Q1.2 Code for LucasKanade()

```
# We recommend using this function, but you can explore other methods as well (e.g.,
ndimage.shift).
from scipy.interpolate import RectBivariateSpline

# The function below could be useful as well :)
# from numpy.linalg import lstsq

def LucasKanade(It, It1, rect, threshold, num_iters, p0=np.zeros(2)):

    # Initialize p to p0. Don't remove these lines.
    p = p0
    delta_p = np.ones((1,2))

    # ---------------------------------------------------------------------------------------
    # TODO: Add your LK implementation here:
    x1,y1,x2,y2=rect
    w,h=int(x2-x1),int(y2-y1)
    rbs0=RectBivariateSpline(np.arange(0,It.shape[0],1),np.arange(0,It.shape[1],1),It)
    rbs1=RectBivariateSpline(np.arange(0,It1.shape[0],1),np.arange(0,It1.shape[1],1),It1)
    change=1
    num_iters=int(num_iters)
    X,Y=np.mgrid[x1:x2+1:w*1j,y1:y2+1:h*1j]
    for ite in range(int(num_iters)):
        if change>threshold:
            px1=rbs1.ev(p[1]+Y,p[0]+X,dy=1).flatten()
            py1=rbs1.ev(p[1]+Y,p[0]+X,dx=1).flatten()
            A=np.zeros((w*h,2*w*h))
            for i in range(w*h):
                A[i,2*i],A[i,2*i+1]=px1[i],py1[i]
delta_p,_,_,_=np.linalg.lstsq(np.matmul(A,mb.repmat(np.eye(2),w*h,1)),np.reshape(rbs0.ev(Y,X).
flatten()-rbs1.ev(p[1]+Y,p[0]+X).flatten(),(w*h,1)),rcond=None)
            change=np.linalg.norm(delta_p)
            p=(p+delta_p.T).ravel()
        else:
            break
# ---------------------------------------------------------------------------------------
    return p
```

1.3 Tracking with *naive* template update (10 points)

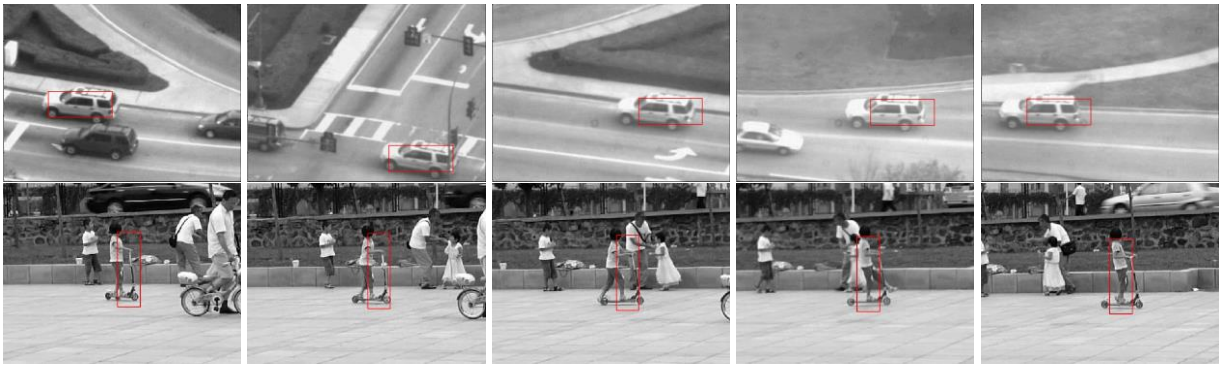


Figure 1.1: Lucas-Kanade Tracking with One Single Template

You will now test your Lucas-Kanade tracker on the `carseq.npy` and `girlseq.npy` sequences.

For this part, you have to implement the following function,

```
rects = TrackSequence(seq, rect, num_iters, threshold)
```

which receives a sequence of frames, the initial coordinates for the template to be tracked, and the number of iterations and threshold for running the LK optimization. The function will use the LK tracker you implemented in **Q1.2** to estimate the motion of the template at each frame. Finally, it must return a matrix containing the rectangle coordinates of the tracked template at each frame.

Once you implement the above function, for the **car sequence** you will have to:

1. Load the `carseq.npy` frame sequence. This sequence has a shape (H, W, N_c).
2. Run `TrackSequence` to track the car (see the top row in fig. 1.1). The `rects` matrix should have a shape $N_c \times 4$. At frame 1, the car is located at coordinates $\text{rect} = [59, 116, 145, 151]^T$.
3. Provide tracking visualizations for frames [1, 80, 160, 280, 410]. These frames are not the same as those in fig. 1.1, but the visualizations should look be similar. Please use the box below to add your visualizations.
4. Save the resulting `rects` to a file called `carseqrects.npy`. You will submit it to Gradescope.

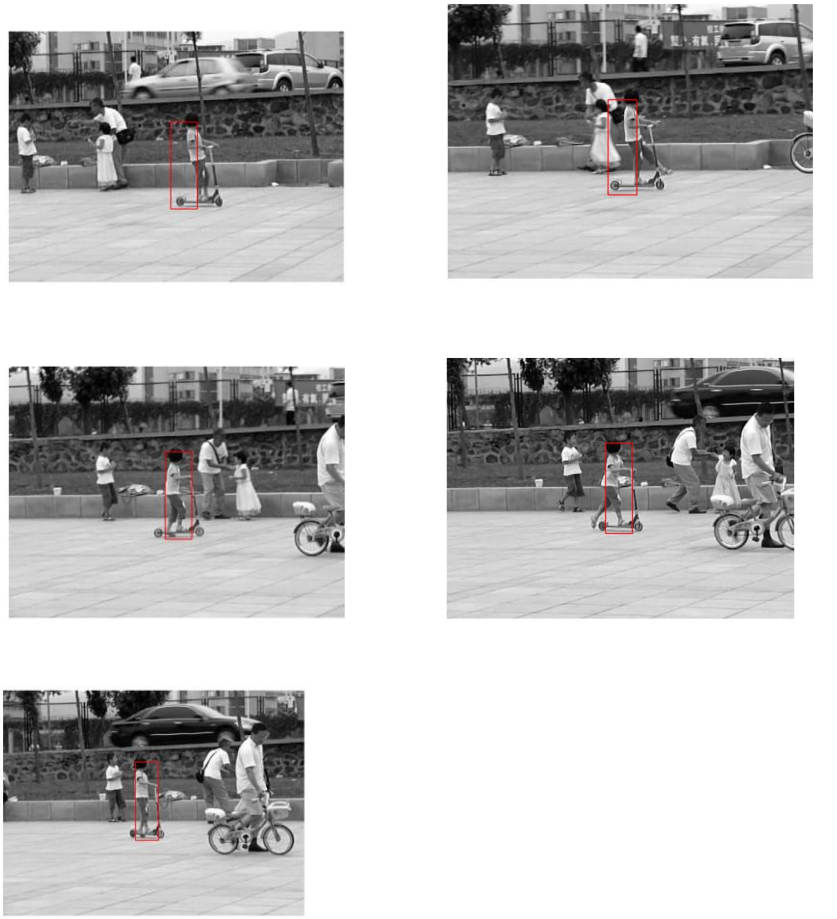
```
def TrackSequenceWithTemplateCorrection(seq, rect_0, num_iters, lk_threshold, drift_threshold):
    """
    H, W, N = seq.shape
    rect = np.copy(rect_0)
    rects_wtcr =[rect_0]
    p0 = np.zeros(2)
    It = seq[:,:,0]
    for frame in tqdm(range(N-1)):
        p=LucasKanade(It,seq[:,:,frame+1],rect,lk_threshold,num_iters,p0)
        p1=p+[rect[0]-rect_0[0],rect[1]-rect_0[1]]
        ps=LucasKanade(seq[:,:,0],seq[:,:,frame+1],rect_0,threshold,num_iters,p1)
        if np.linalg.norm(p1-ps)<threshold:
            p2=(ps-[rect[0]-rect_0[0],rect[1]-rect_0[1]])
            rect=[p2[0]+rect[0],p2[1]+rect[1],p2[0]+rect[2],p2[1]+rect[3]]
            p0=np.zeros(2)
            It=seq[:,:,frame+1]
            rects_wtcr.append(rect)
        else:
            rects_wtcr.append([rect[0]+p[0],rect[1]+p[1],rect[2]+p[0],rect[3]+p[1]])
            p0 = p
    # ----------------------------------------------------------------------------------------------
    rects_wtcr = np.array(rects_wtcr)
    # Just a sanity check
    assert rects_wtcr.shape == (N, 4), f"Your output sequence {rects_wtcr.shape} is not {N}x{4}"
    return rects_wtcr
```

Q1.3 Car sequence visualizations

For the **girl sequence** you will have to:

1. Load the `girlseq.npy` frame sequence. This sequence has a shape (H, W, N_g).
2. Run the `TrackSequence` to track the girl (see bottom row in fig. 1.1). The `rects` matrix should have a shape $N_g \times 4$. At frame 1, the girl is located at coordinates $\text{rect} = [280, 152, 330, 318]^T$.
3. Provide tracking visualizations for frames [1, 15, 35, 65, 85]. These frames are not the same as those in fig. 1.1, but the visualizations should look similar. Please use the box below to add your visualizations.
4. Save the resulting `rects` to a file called `girlseqrects.npy`. You will submit it to Gradescope.

Q1.3 Girl sequence visualizations



Notes:

- A frame can be visualized as `plt.imshow(frames[:, :, i]); plt.show()`
- This equation might be useful for updating the coordinates of the rectangle $T_{t+1}(\mathbf{x}) = I_t(W(\mathbf{x}; \mathbf{p}_t))$.

- We provided initial values for the threshold and the number of iterations for the tracker, but you are encouraged to play with the parameters defined in the scripts and report the best results.

1.4 Tracking with template correction (20 points)

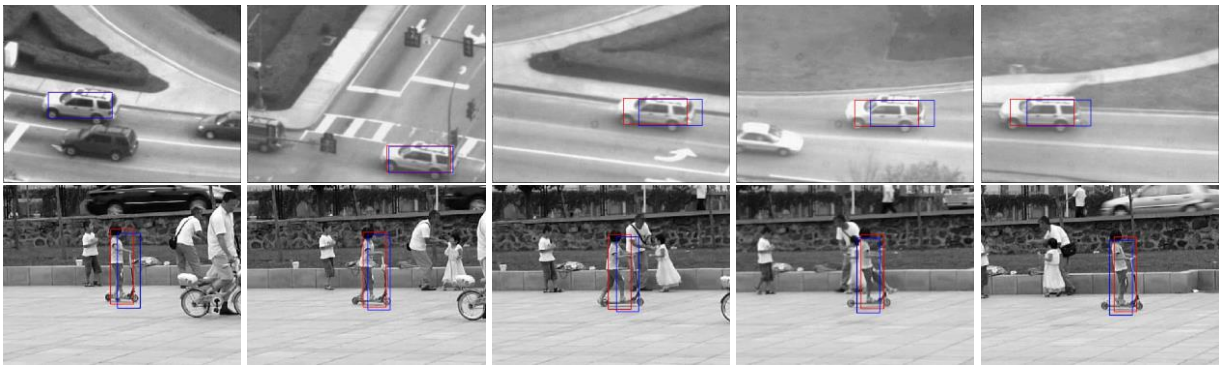


Figure 1.2: Lucas-Kanade Tracking with Template Correction

As you might have noticed, the image content we are tracking in the first frame differs from that in the last frame. This issue is known as *template drifting*, and it is due to error accumulation that stems from doing a *naive* update of the template. There are several template update strategies for mitigating this drifting problem. The one we will explore here is explained in the paper below,

- Ian Matthews, et al. *The Template Update Problem*. Proceedings of British Machine Vision Conference (BMVC '03), 2003. [link]

This method proposes an extension to the *naive* update. Roughly, the idea is to update the template at each step, as before, but it must be re-aligned to the initial template, T_0, to correct for drift [3].

For this question, you will follow a similar process as in **Q1.3** to test, both, the **car** and **girl** sequences. Except, now, we additionally ask you to implement the template update strategy from the previous paper. Specifically, you will need to implement **Strategy 3: Template Update with Drift Correction** which follows the update below,

$$\begin{aligned} &\textbf{if} \quad \|\mathbf{p}_t^* - \mathbf{p}_t\| \leq \epsilon \quad \textbf{then} \quad T_{t+1}(\mathbf{x}) = \boldsymbol{I}_t(W(\mathbf{x}; \mathbf{p}^*)) \\ &\textbf{else} \quad T_{t+1}(\mathbf{x}) = T_t(\mathbf{x}) \end{aligned}$$

An example of LK with template correction is given in fig. 1.2. The blue rectangles are results obtained with the baseline tracker (**Q1.3**), the red ones are results obtained with the tracker with drift correction (**Q1.4**).

Before implementing the drift correction, please read **section 2.1** and **section 2.3** of the paper. Note that you **do not** need to modify your Lucas-Kanade implementation. Once you're done reading these sections, implement the function,

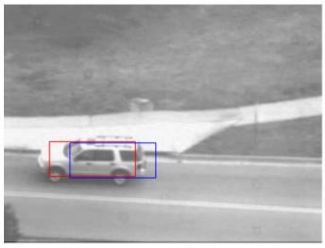
```
rects = TrackSequenceWithTemplateCorrection(seq, rect, num_iters, lk_threshold,
                                            drift_threshold)
```

which, as before, receives a frame sequence, the initial coordinates of the object of interest, the number of iterations, and the threshold for running the optimization, and now, it also receives **the drift threshold parameter for the template update.**

Then, follow a similar procedure as in **Q1.3** to test and visualize your results, *i.e.* follow this to-do list for the **car sequence**,

1. Load the `carseq.npy` frame sequence.
2. Run `TrackSequenceWithTemplateCorrection` to track the car.
3. Provide tracking visualizations for frames [1, 80, 160, 280, 410]. Here you will provide visualizations **with** and **without** template correction in red and blue, respectively as in fig. 1.2.
4. Save the result to a file called `carseqrects-wtcr.npy`. You will submit it to Gradescope.

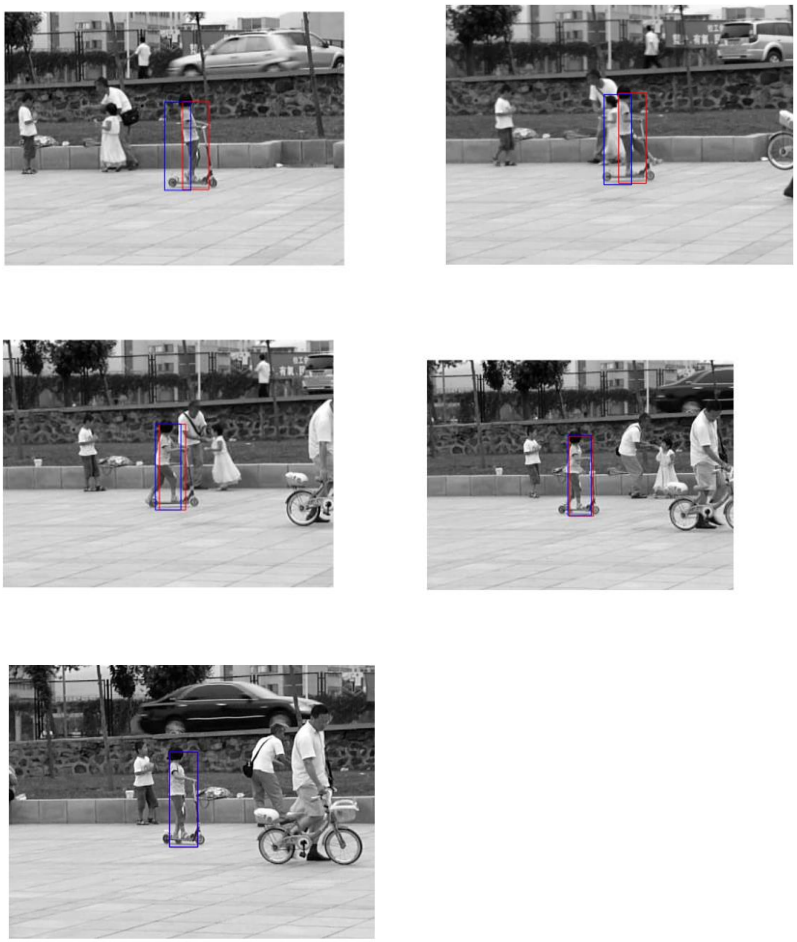
Q1.4 Car Sequence with and without template correction



And follow this to-do list for the **girl sequence**,

1. Load the `girlseq.npy` frame sequence.
2. Run `TrackSequenceWithTemplateCorrection` to track the girl.
3. Provide tracking visualizations for frames [1, 15, 35, 65, 85]. Here you will provide visualizations **with** and **without** template correction in red and blue, respectively as in fig. 1.2.
4. Save the result to a file called `girlseqrects-wtcr.npy`. You will submit it to Gradescope.

Q1.4 Girl sequence with and without template correction



Notes:

- Again, we provide you with initial parameters, but you are encouraged to play with them to see how each parameter affects the tracking results and report your best ones.

2 Affine Motion Subtraction (35 total points)

In the first section of this homework, we assumed the motion is limited to pure translation and a single template. In this section, you will now implement a tracker for **affine motion**. For implementing such tracker, you will be working on two main parts: 1) estimating the dominant affine motion in subsequent images (section 2.1) ; and 2) identifying pixels corresponding to moving objects in the scene (section 2.2). For testing it, as before, you will be asked to visualize the results of your implementation (section 2.3).

Coding questions for **part 2** should be implemented in `LucasKanadeAffine.ipynb`. We provide starter code for each question, as well as default parameters. You will test your tracker, on an ant sequence, `antseq.npy` (top row in fig. 2.1), and an aerial sequence of moving vehicles, `aerialseq.npy` (bottom row in fig. 2.1), both of which you can find in the `data` folder.

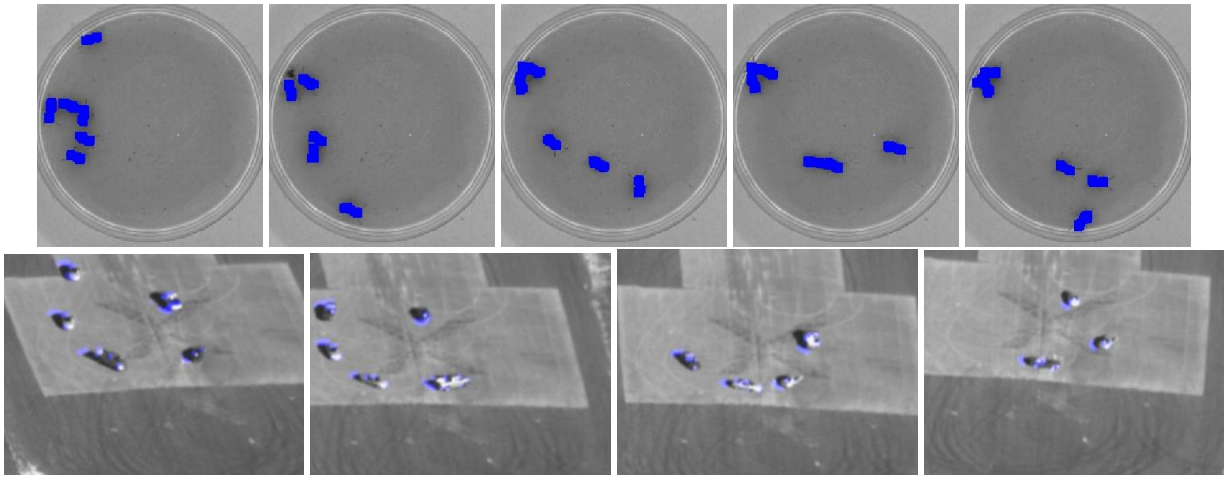


Figure 2.1: Lucas-Kanade Tracking with Motion Detection

2.1 Dominant Motion Estimation (15 points)

In this section, you will implement a tracker for affine motion using a planar affine warp function. To estimate the dominant motion, the entire image $\boldsymbol{I}_t$ will serve as the template to be tracked in image $\boldsymbol{I}_{t+1}$, that is, $\boldsymbol{I}_{t+1}$ is assumed to be approximately an affine warped version of $\boldsymbol{I}_t$. This approach is reasonable under the assumption that a majority of the pixels correspond to stationary objects in the scene whose depth variation is small relative to their distance from the camera.

Let us now define an **affine** warp function as,

$$\mathbf{x}' = W(\mathbf{x}; \mathbf{p}) = \begin{bmatrix} 1+p_1 & p_2 \\ p_4 & 1+p_5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} p_3 \\ p_6 \end{bmatrix} . \tag{2.1}$$

As described in [2], one can represent this affine warp in homogeneous coordinates as,

$$\tilde{\mathbf{x}}' = \mathbf{M}\tilde{\mathbf{x}} \tag{2.2}$$

where

$$\mathbf{M} = \begin{bmatrix} 1+p_1 & p_2 & p_3 \\ p_4 & 1+p_5 & p_6 \\ 0 & 0 & 1 \end{bmatrix} . \tag{2.3}$$

where $\mathbf{M}$ will differ between successive image pairs.

Similar as before, the algorithm starts with an initial guess of $\mathbf{p} = [0, 0, 0, 0, 0, 0]^T$, *i.e.* $\mathbf{M} = \mathrm{I}$. To determine $\Delta\mathbf{p}$, you will need to iteratively solve a least-squares such that $\mathbf{p} \rightarrow \mathbf{p} + \Delta\mathbf{p}$ at each iteration.

Write the function,

```
M = LucasKanadeAffine(It, It1, num_iters, threshold)
```

where the input parameters are similar to those of `LucasKanade` in **Q1.2**, but note that we are **not** using `rect` since we will use the entire image as the template. The function should now return a 3×3 affine transformation matrix `M`.

Notes:

- `LucasKanadeAffine` should be relatively similar to `LucasKanade`. In **Q1.2** the template to be tracked is usually small, compared to tracking the whole image. For this part, image I_t will almost always not be fully contained in the warped version I_{t+1}. Therefore, the matrix of image derivatives, $\mathbf{A}$, and the temporal derivatives, ∂I_t, must be computed only on the pixels lying in the region common to I_t and the warped version of I_{t+1}.

Include the code you wrote for this part in the box below:

Q2.1 Code for LucasKanadeAffine

```
# We recommend using this function, but you can explore other methods as well (e.g.,
ndimage.shift).
from scipy.interpolate import RectBivariateSpline

# The function below could be useful as well :)
# from numpy.linalg import lstsq

def LucasKanadeAffine(It, It1, threshold, num_iters):
  # Initial M
   # M = np.eye(3)
   M = np.array([[1.0, 0.0, 0.0], [0.0, 1.0, 0.0]])

   # ---------------------------------------------------------------------------------------
   # TODO: Add your LK implementation here:
   p=[1,0,0,0,1,0]
   it1_rbs=RectBivariateSpline(np.arange(It1.shape[0]), np.arange(It1.shape[1]), It1)
   dp=np.ones((1,5))
   while np.linalg.norm(dp)>=threshold:
      coord_x,coord_y=np.meshgrid(np.arange(0,It.shape[1]), np.arange(0, It.shape[0]))
      coord_xn = p[0]*coord_x+p[1]*coord_x+p[2]
      coord_yn = p[3]*coord_y+p[4]*coord_y+p[5]
      coord_x,coord_y=coord_x.flatten(),coord_y.flatten()
      gintx,ginty = it1_rbs.ev(coord_yn,coord_xn, dx=0, dy=1).flatten(),it1_rbs.ev(coord_yn,
coord_xn, dx=1, dy=0).flatten()

A=np.column_stack([np.multiply(gintx,coord_x.flatten()),np.multiply(gintx,coord_y.flatten()),gin
tx,np.multiply(ginty,coord_x.flatten()),np.multiply(ginty,coord_y.flatten()),ginty])
      Iint = it1_rbs.ev(coord_yn,coord_xn)
      b=It.flatten()-Iint.flatten()
      dp,_,_,_ = np.linalg.lstsq(np.dot(np.transpose(A),A),np.dot(np.transpose(A), b),rcond=None)
      p+=dp.flatten()
   M = np.reshape(p, (2, 3))
   # ---------------------------------------------------------------------------------------
   return M
```

2.2 Moving Object Detection (10 points)

Once you're able to compute the affine warp $\mathbf{M}$ between the image pair $\boldsymbol{I}_t$ and $\boldsymbol{I}_{t+1}$, you have to determine the pixels corresponding to moving objects. One naive way to do so is as follows:

1. Warp the image $\boldsymbol{I}_t$ using $\mathbf{M}$ so that it is registered to $\boldsymbol{I}_{t+1}$. To do this, you may find the functions `scipy.ndimage.affine_transform` or `cv2.warpAffine` useful. Please carefully read their corresponding documentation on whether $\mathbf{M}$ or $\mathbf{M}^{-1}$ needs to be passed to the function.
2. Subtract the warped image from $\boldsymbol{I}_{t+1}$; the locations where the absolute difference exceeds a tolerance can then be declared as corresponding to the locations of moving objects. To obtain better results, you might find the following functions useful: `scipy.morphology.binary_erosion`, and `scipy.morphology.binary_dilation`.

Write the following function,

```
mask = SubtractDominantMotion(It, It1, num_iters, threshold, tolerance)
```

which receives the image pair `It` and `It1`, the number of iterations and threshold parameters for running the LK optimization, and the **tolerance to determine the moving pixels**. The function must return a `mask` which is a binary image specifying which pixels correspond to moving objects. Note that you should use `LucasKanadeAffine` within this function to derive the transformation matrix $\mathbf{M}$, and produce the according binary mask.

Include the code you wrote for this part in the box below:

Q2.2 Code for SubtractDominantMotion

```
# These functions could be useful for your implementation.
from scipy.ndimage import binary_erosion, binary_dilation, affine_transform
import cv2

def SubtractDominantMotion(It, It1, num_iters, threshold, tolerance):

    mask = np.zeros(It1.shape, dtype=bool)

    # ----------------------------------------------------------------------------------------
    # TODO: Add your code here:
    M = LucasKanadeAffine(It,It1, threshold, num_iters)
    if M.shape[0] < 3:
        M = np.vstack((M, np.array([[0, 0, 1]])))
    M = np.linalg.inv(M)
    warp_image1 = scipy.ndimage.affine_transform(It,M[0:2,0:2],offset = M[0:2,2],output_shape =
It1.shape)
    diff=abs(warp_image1-It1)
    mask[abs(warp_image1-It1)>tolerance]=1
    mask[warp_image1==0]=0
    # ----------------------------------------------------------------------------------------
    return mask
```

2.3 Tracking with affine motion (10 points)

Similar to **Q1.3** and **Q1.4**, you will now test your implementation of the Lucas-Kanade tracker with affine motion on the `antseq.npy` and `aerialseq.npy`.

Implement the function,

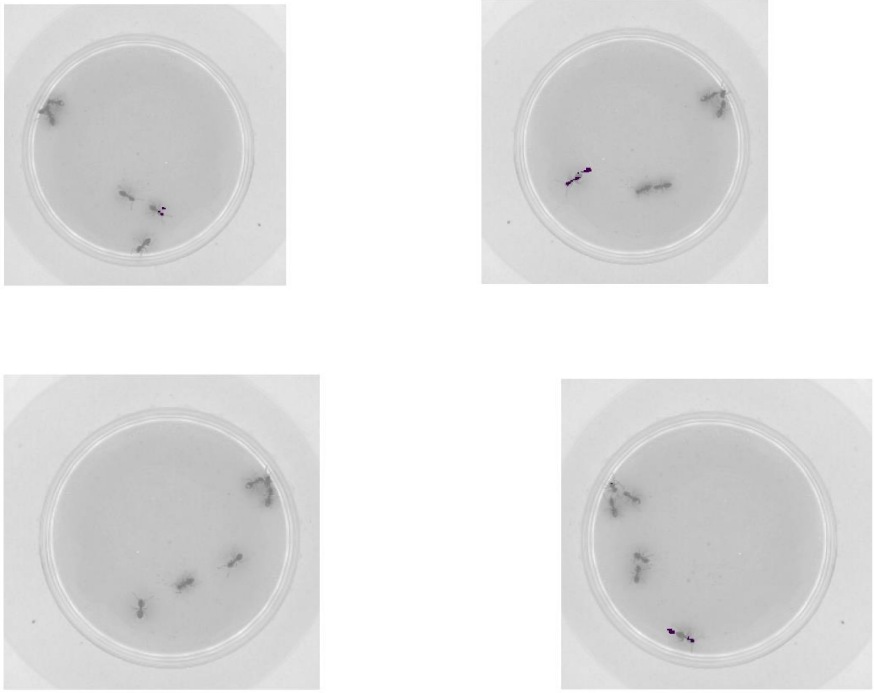
```
mask = TrackSequenceAffineMotion(seq, num_iters, threshold, tolerance)
```

which receives a sequence of frames, the number of iterations and threshold for running the optimization and the tolerance for the motion subtraction. The function must return `masks`, a matrix which contains the binary outputs from the motion subtraction method you implemented in the previous section.

Then, follow a similar procedure as in **Q1.3** to test and visualize your results, *i.e.* follow this to-do list for the **ant sequence**,

1. Loads the `antseq.npy` frame sequence.
2. Runs `TrackSequenceAffineMotion` to track the ants.
3. Provides visualizations for frames [30, 60, 90, 120], overlaying the corresponding binary masks similar to those in fig. 2.1.
4. You will **not** upload the `masks` to Gradescope.
5. Report your run-time performance.

Q2.3 Ant sequence visualizations



Q2.3 Ant sequence run-time performance

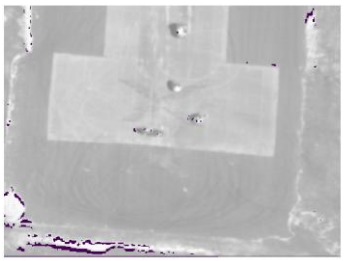
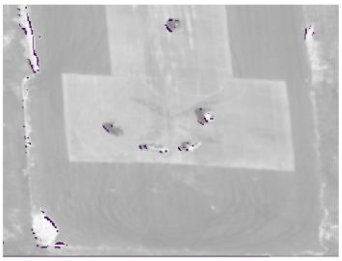
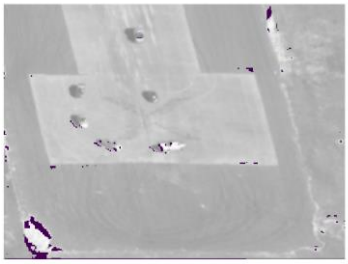
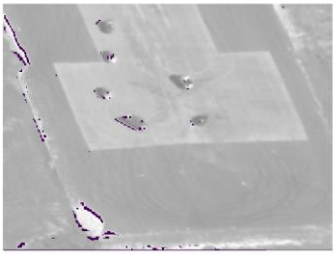
```
159707 function calls (158206 primitive calls) in 14.250 seconds
```

And follow this to-do list for the **aerial sequence**,

1. Loads the `aerial.npy` frame sequence.
2. Runs `TrackSequenceAffineMotion` to track the cars.
3. Provides visualizations for frames [30, 60, 90, 120], overlaying the corresponding binary masks similar to those in fig. 2.1.
4. You will **not** upload the `masks` to Gradescope.

5. Report your run-time performance.

Q2.3 Aerial sequence Visualizations

Q2.3 Aerial sequence run-time performance

```
349387 function calls (345357 primitive calls) in 57.818 seconds
```

Notes:

1. The **ant sequence** involves little camera movement and can help you debug your mask generation procedure.
2. Feel free to visualize the motion detection performance in a way that you would prefer, but please make sure it can be visually inspected without undue effort.

3 Efficient Tracking (15 total points)

In this section, you will explore Lucas-Kanade with inverse composition for efficient tracking.

Coding questions for **part 3** should be implemented in `LucasKanadeEfficient.ipynb`. Again, we provide starter code for each question, as well as default parameters. You will test your tracker, on `antseq.npy` and `aerialseq.npy`.

3.1 Inverse Composition (10 points)

The inverse compositional extension of the Lucas-Kanade algorithm [2] has been used in literature to great effect for the task of efficient tracking. When utilized within tracking it attempts to linearize the current frame as:

$$I_t(W(\mathbf{x}; \mathbf{0} + \Delta\mathbf{p}) \approx I_t(\mathbf{x}) + \frac{\partial I_t(\mathbf{x})}{\partial \mathbf{x}^T} \frac{\partial W(\mathbf{x}; \mathbf{0})}{\partial \mathbf{p}^T} \Delta\mathbf{p} \quad . \tag{3.1}$$

In a similar manner to the conventional Lucas-Kanade algorithm one can incorporate these linearized approximations into a vectorized form such that,

$$\arg\min_{\Delta\mathbf{p}} ||\mathbf{A}'\Delta\mathbf{p} - \mathbf{b}'||^2_2 \tag{3.2}$$

for the specific case of an affine warp where we can recover $\mathbf{p}$ from $\mathbf{M}$ and $\Delta\mathbf{p}$ from $\Delta\mathbf{M}$. This results in the update $\mathbf{M} = \mathbf{M}(\Delta\mathbf{M})^{-1}$.
Here,

$$\Delta\mathbf{M} = \begin{bmatrix} 1 + \Delta p_1 & \Delta p_2 & \Delta p_3 \\ \Delta p_4 & 1 + \Delta p_5 & \Delta p_6 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.3}$$

With this in mind, write the function,

```
M = InverseCompositionAffine(It, It1, num_iters, threshold)
```

which re-implements function `LucasKanadeAffine` from **Q2.1**, but now using the aforesaid inverse compositional method.

Include the code you wrote for this part in the box below:

Q3.1 Code for InverseCompositionAffine

```
from scipy.ndimage import affine_transform
import cv2
import scipy.ndimage
from scipy.interpolate import RectBivariateSpline
from scipy.ndimage.morphology import binary_opening, binary_closing
from scipy.ndimage import affine_transform

def InverseCompositionAffine(It, It1, threshold, num_iters):
    M = np.array([[1.0, 0.0, 0.0], [0.0, 1.0, 0.0]])
    # TODO: Add your Efficient LK implementation here:
    p=[1,0,0,0,1,0]
    dp=np.ones((1,5))
    it1_rbs=RectBivariateSpline(np.arange(It.shape[0]),np.arange(It.shape[1]),It1)
    it_rbs=RectBivariateSpline(np.arange(It.shape[0]),np.arange(It.shape[1]),It)
    coord_x,coord_y= np.meshgrid(np.arange(0,It.shape[1]),np.arange(0,It.shape[0]))
    gintx,ginty =
it_rbs.ev(coord_y,coord_x,dx=0,dy=1).flatten(),it_rbs.ev(coord_y,coord_x,dx=1,dy=0).flatten()
    A=np.column_stack([np.multiply(gintx,coord_x.flatten()),np.multiply(gintx,
coord_y.flatten()),gintx,np.multiply(ginty, coord_x.flatten()),np.multiply(ginty,
coord_y.flatten()),ginty])
    while np.linalg.norm(dp)>=threshold:
        coord_xn = p[0]*coord_x+p[1]*coord_y+p[2]
        coord_yn = p[3]*coord_x+p[4]*coord_y+p[5]
        valid=(coord_xn > 0)&(coord_xn<It1.shape[1])&(coord_yn>0)&(coord_yn<It1.shape[0])
        coord_xn,coord_yn=coord_xn[valid],coord_yn[valid]
        Iint=it1_rbs.ev(coord_yn,coord_xn)
        A1=A[valid.flatten()]
        b=Iint.flatten()-It[valid].flatten()
        b1=np.dot(A1.T, b)
        dp,_,_,_ = np.linalg.lstsq(np.dot(A1.T,A1),b1,rcond=None)
        M=np.vstack((np.reshape(p, (2, 3)),np.array([[0, 0, 1]])))
        dm=np.vstack((np.reshape(dp, (2, 3)), np.array([[0, 0, 1]])))
        dm[0,0]+=1
        dm[1,1]+=1
        M=np.dot(M,np.linalg.inv(dm))
        p=M[:2, :].flatten()

    M=M[:2, :]
    # ---------------------------------------------------------------------------------------
    return M
```

Notes:

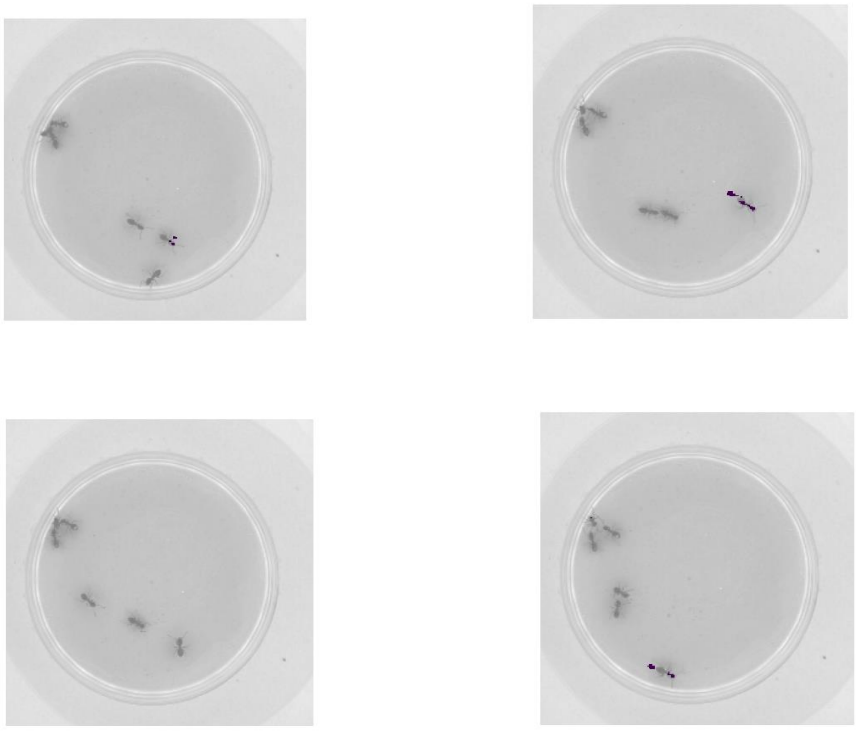
1. The notation $\mathbf{M}(\Delta\mathbf{M})^{-1}$ corresponds to $W(W(\mathbf{x};\Delta\mathbf{p})^{-1};\mathbf{p})$ in Section 2.2 in [1].

3.2 Tracking with Inverse Composition (5 points)

You will now re-use your `SubtractDominantMotion` and `TrackSequenceAffineMotion` implemented in **Q2.2** and **Q2.3** to provide the visualizations for the ant and aerial sequences.

Thus, for the **ant sequence** follow the same to-do list as in **Q2.3**. When reporting your run-time performance, please also include the percentage gain with respect to **Q2.3**.

Q3.2 Ant sequence visualizations

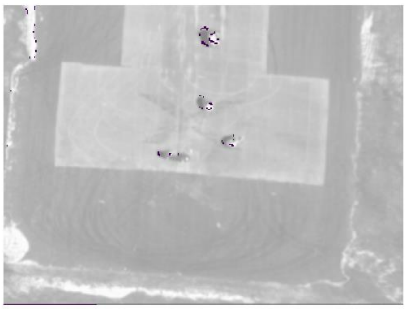


For the **aerial sequence** follow the same to-do list as in **Q2.3**. When reporting your run-time performance, please also include the percentage gain with respect to **Q2.3**.

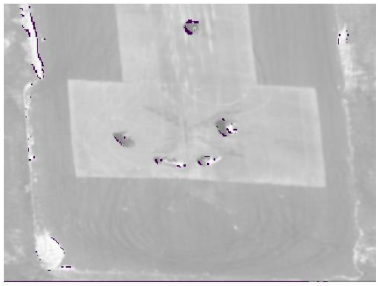
Q3.2 Ant sequence run-time performance and performance gain

```
138059 function calls (135778 primitive calls) in 10.445 seconds
So, performance gain = (14.25-10.445)/10.445*100 = 36.42%
```

Q3.2 Aerial sequence visualizations

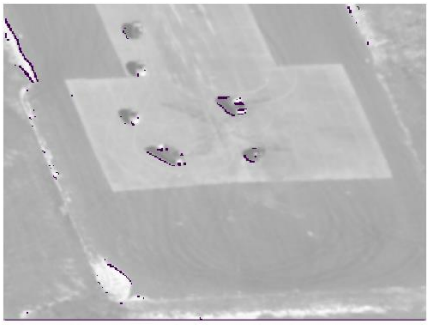


Frame 30



Frame 60

Frame 90



Frame 120

Q3.2 Aerial sequence run-time performance and performance gain

```
326307 function calls (319821 primitive calls) in 27.687 seconds
```

Therefore, performance gain = (57.818-27.687)/57.818*100 = 52.11%

Finally, in your own words, please describe briefly why the inverse compositional approach is more computationally efficient than the classical approach.

Q3.2
Answer

The inverse affine compositional approach is more computationally efficient than the classical approach because it reduces the number of computations needed to warp an image. In the classical approach, the warp function is computed for every pixel in the image, which can be computationally expensive for large images or complex warp functions.

In contrast, the inverse affine compositional approach computes the warp function once and applies it to the entire image. This approach is possible because the warp function is linear and can be expressed as an affine matrix. By computing the inverse of the affine matrix, we can apply the warp function to the entire image with a single matrix multiplication.

Additionally, the inverse affine compositional approach allows for efficient computation of the gradient of the warped image, which is important for tasks such as image alignment or feature tracking.

4 Extra credit

Find a 10s video clip of your choice and run Lucas-Kanade tracking on a salient foreground object in this video. Needless to say, the object you track should undergo considerable amount of motion in the scene and should not be static. It is even better if this object undergoes an occlusion and your algorithm is able to track it across this occlusion. To report your results in the write-up, capture 6 frames of the tracked video at 0s, 2s, 4s, 6s, 8s, 10s and include these.

Visualizations for the corresponding frames

Frame at 0s

Frame at 20s

Frame at 40s

Frame at 60s

Frame at 80s

Frame at 100s

Explain any changes you had to make in the algorithm to get it to work on this video.

Answer

Here, I've taken a video of myself moving a pendrive infront of the camera and the objective here was to use LK Tracking to track the pendrive through the sequence of frames of the video.

To the algorithm, the entire logic remains the same. The slicing operation on each frame changes from seq[:,:,i] to seq[i,:,:] and number of frames here is referred to as seq.shape[0]

To make the tracking less computational intensive, I used a video editor to change the aspect ratio by decreasing the number of pixels on both the axes.

The initial bounding box had to be changed that could locate the pendrive in the initial frame and it was rect = [15,40,75,70]

5 Deliverables

The assignment (code and writeup) should be submitted to Gradescope. The write-up should be submitted to Gradescope named <AndrewId> hw2.pdf. The code should be submitted as a zip named <AndrewId> hw2.zip to Gradescope. The zip when uncompressed should produce the following files.

- `LucasKanade.ipynb`
- `LukasKanadeAffine.ipynb`
- `LukasKanadeEfficient.ipynb`
- (Optional) `ExtracCredit.ipynb`
- `carseqrects.npy`
- `carseqrects-wtcr.npy`
- `girlseqrects.npy`
- `girlseqrects-wtcr.npy`

***Do not include the data directory in your submission.**

6 Frequently Asked Questions (FAQs)

Q1: Why do we need to use `ndimage.shift` or `RectBivariateSpline` for moving the rectangle template?

A1: When moving the rectangle template with $\Delta\mathbf{p}$, you can either move the points inside the template or move the image in the opposite direction. If you choose to move the points, the new points can have fractional coordinates, so you need to use `RectBivariateSpline` to sample the image intensity at those fractional coordinates. If you instead choose to move the image with `ndimage.shift`, you don't need to move the points and you can sample the image intensity at those points directly. The first approach could be faster since it does not require moving the entire image.

Q2: What's the right way of computing the image gradients $I_x(\mathbf{x})$ and $I_y(\mathbf{x})$. Should I first sample the image intensities $I(\mathbf{x})$ at $\mathbf{x}$ and then compute the image gradients $I_x(\mathbf{x})$ and $I_y(\mathbf{x})$ with $I(\mathbf{x})$? Or should I first compute the entire image gradients $\boldsymbol{I}_x$ and $\boldsymbol{I}_y$ and sample them at $\mathbf{x}$?

A2: The second approach is the correct one.

Q3: Can I use pseudo-inverse for the least-squared problem $\arg\min_{\Delta\mathbf{p}} ||\mathbf{A}\Delta\mathbf{p} - \mathbf{b}||_2^2$?

A3: Yes, the pseudo-inverse solution of $\mathbf{A}\Delta\mathbf{p} = \mathbf{b}$ is also $\Delta\mathbf{p} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$ when $\mathbf{A}$ has full column ranks, i.e., the linear system is overdetermined.

Q4: For inverse compositional Lucas-Kanade, how should I deal with points outside out the image?

A4: Since the Hessian in inverse compositional Lucas Kanade is precomputed, we cannot simply remove points when they are outside the image since it can result in dimension mismatch. However, we can set the error $I_{t+1}(W(\mathbf{x};\mathbf{p})) - I_t(\mathbf{x})$ to 0 for $\mathbf{x}$ outside the image.

Q5: How to find pixels common to both It1 and It?

A5: If the coordinates of warped It1 is within the range of It.shape, then we consider the pixel lies in the common region.

References

[1] Simon Baker, Ralph Gross, Iain Matthews, and Takahiro Ishikawa. Lucas-kanade 20 years on: A unifying framework: Part 2. Technical Report CMU-RI-TR-03-01, Carnegie Mellon University, Pittsburgh, PA, February 2003.

[2] Simon Baker and Iain Matthews. Lucas-kanade 20 years on: A unifying framework part 1: The quantity approximated, the warp update rule, and the gradient descent approximation. *International Journal of Computer Vision - IJCV*, 01 2004.

[3] Iain Matthews, Takahiro Ishikawa, and Simon Baker. The template update problem. In *Proceedings of British Machine Vision Conference (BMVC '03)*, September 2003.