EE 046746 - Technion - Computer Vision

Homework 4 - Homographies

Updated 21.6.2020

Due Date: 02.07.2020

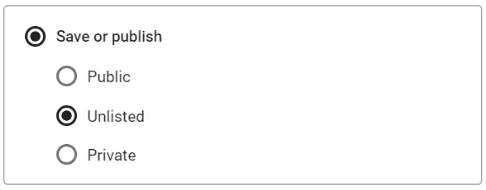
Submission Guidelines

READ THIS CAREFULLY

- Submission only in **pairs**.
- **No handwritten submissions**.
- You can choose your working environment:
  - You can work in a `Jupyter Notebook` , locally with [Anaconda (https://www.anaconda.com/distribution/)](https://www.anaconda.com/distribution/) or online on [Google Colab (https://colab.research.google.com/)](https://colab.research.google.com/)
    - Colab also supports running code on GPU, so if you don't have one, Colab is the way to go. To enable GPU on Colab, in the menu: `Runtime` $\rightarrow$ `Change Runtime Type` $\rightarrow$ `GPU` .
  - You can work in a Python IDE such as [PyCharm (https://www.jetbrains.com/pycharm/)](https://www.jetbrains.com/pycharm/) or [Visual Studio Code (https://code.visualstudio.com/)](https://code.visualstudio.com/).
    - Both also allow opening/editing Jupyter Notebooks.
- You should submit two **separated** files:
  - A compressed `.zip` file, with the name: `ee046746_hw4_id1_id2.zip` which contains:
    - A folder named `code` with all the code files inside (`.py` or `.ipynb` ONLY!), and all the files required for the code to run (your own images/videos) inside `my_data` folder.
      - **The code should run both on CPU and GPU without manual modifications**, require no special preparation and run on every computer.
    - A folder named `output` with all the output files that are not required to run the code. This includes videos and visualizations that are not included in your PDF report.
  - A report file (visualizations, discussing the results and answering the questions) in a `.pdf` format, with the name `ee046746_hw4_id1_id2.pdf` .
    - Be precise, we expect on point answers. **But don't be afraid to explain you statements (actually, we expect you to).**
      - Even if the instructions says "Show...", you still need to explain what are you showing and what can be seen.
  - No other file-types (`.docx` , `.html` , ...) will be accepted.
- Submission on the course website (Moodle).

Video Submission

- In this exercise you are going to produce a video.
- In addition to submititng this video in the `output` folder, we also ask you to upload it to YouTube and submit the link in a different section on the website (just below the original submission section).
  - This will benefit you when you explain your expertise in Computer Vision (e.g., in a job interview).
  - We also want to make something nice with your works at the end of the course.
- If you don't want to make your video public, you can change its visibility to "Unlisted", and then it can only be accessed via link:



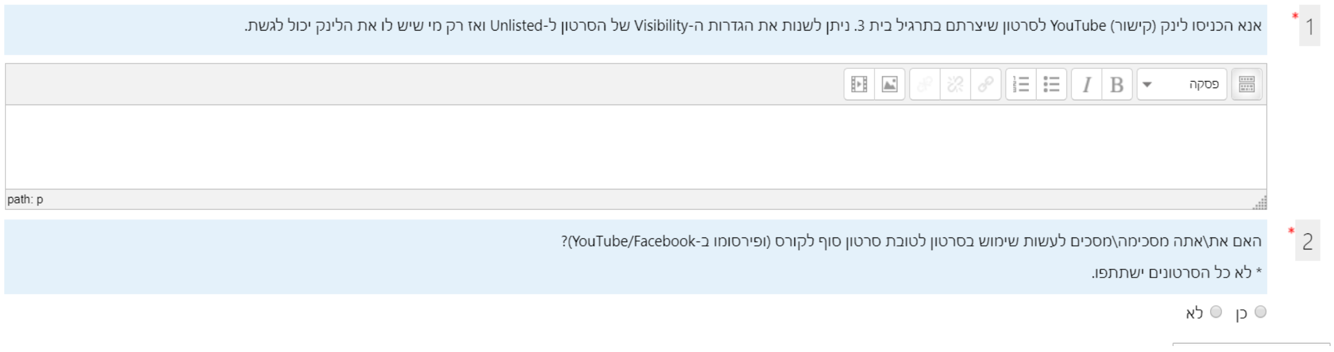
[✓ Save or publish]

Public

[✓ Unlisted]

Private

- Finally, submit the link on the course website and answer if you are okay with making this video public:



1 * אנא הכניסו לינק (קישור) YouTube לסרטון שיצרתם בתרגיל בית 3. ניתן לשנות את הגדרות ה-Visibility של הסרטון ל-Unlisted ואז רק מי שיש לו את הלינק יכול לגשת.

2 * האם את\אתה מסכימה\מסכים לעשות שימוש בסרטון לטובת סרטון סוף לקורס (ופירסומו ב-YouTube/Facebook)?

* לא כל הסרטונים ישתתפו.

כן לא

Python Libraries

- `numpy`
- `matplotlib`
- `opencv` (or `scikit-image`)
- `scikit-learn`
- `scipy`
- `torch` (and `torchvision`)
- Anything else you need (`os`, `pandas`, `csv`, `json`,...)

Important Tips

- You can resize the images to a lower resolution if the computation takes too long.
- You can parameters to the functions if you need.
- In part 3, you can use the OpenCV warping functions, or any other OpenCV functions for a faster implementation.

Tasks

- In all tasks, you should document your process and results in a report file (which will be saved as `.pdf`).
- You can reference your code in the report file, but no need for actual code in this file, the code is submitted in a seprate folder as explained above.

Part 1 - Theory

Let's take a look at some simple theory questions. The answers to the below questions should be relatively short, consisting of a few lines of math and text (maybe a diagram if it helps your understanding).

Question 1.1. Suppose two cameras fixate on a point P(see Figure 1) in space such that their principal axes intersect at that point. Show that if the image coordinates are normalized so that the coordinate origin (0,0) coincides with the principal point, the F33 element of the fundamental matrix is zero.

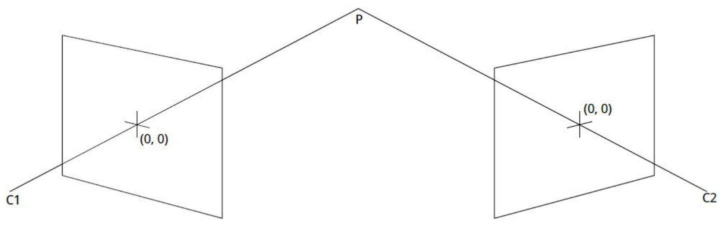


Figure 1: Figure for Q1.1. $C1$ and $C2$ are the optical centers. The principal axes intersect at point $\mathbf{w}$ (P in the figure).

Note, pricipal point is the image center, and pricipal axis can be consider as the Z axis (pointing in the viewing direction).

Question 1.2. Consider the case of two cameras viewing an object such that the second camera differs from the first by a pure translation that is parallel to the x-axis. Show that the epipolar lines in the two cameras are also parallel to the x-axis. Backup your argument with relevant equations.

Question 1.3. Suppose we have an inertial sensor which gives us the accurate positions(Ri and ti, where R is the rotation matrix and t is corresponding translation vector) of the robot at time i. What will be the effective rotation (Rrel) and translation (trel) between two frames at different time stamps? Suppose the camera intrinsics (K) are known, express the essential matrix(E) and the fundamental matrix (F) in terms of K, Rrel and trel.

Part 2 & 3: Introduction

In this homework, we will explore the homography between images based on the locations of the matched features (remember HW1?). Specifically, we will look at the planar homographies. Why is this useful? In many robotics applications, robots must often deal with tabletops, ground, and walls among other flat planar surfaces. When two cameras observe a plane, there exists a relationship between the captured images. This relationship is defined by a 3×3 transformation matrix, called a planar homography. A planar homography allows us to compute how a planar scene would look from a second camera location, given only the first camera image. In fact, we can compute how images of the planes will look from any camera at any location without knowing any internal camera parameters and without actually taking the pictures, all using the planar homography matrix.

Part 2.A - Planar Homographies: Theory warm up

Suppose we have two cameras C_1 and C_2 looking at a common plane Π in 3D space. Any 3D point P on Π generates a projected 2D point located at $p \equiv (u_1, v_1, 1)^T$ on the first camera C_1 and $q \equiv (u_2, v_2, 1)^T$ on the second camera C_2. Since P is confined to the plane Π, we expect that there is a relationship between p and q. In particular, there exists a common 3×3 matrix H, so that for any P, the following conditions holds:

$$(1)\ p \equiv Hq$$

We call this relationship *'planar homography'*. Recall that both p and q are in homogeneous coordinates and the equality $\equiv$ means p is proportional to Hq (recall homogeneous coordinates). It turns out this relationship is also true for cameras that are related by pure rotation without the planar constraint.

Q2.0:

We have a set of points $p = \{p_1, p_2, \dots, p_N\}$ in an image taken by camera C_1 and corresponding points $q = \{q_1, q_2, \dots, q_N\}$ in an image taken by C_2. Suppose we know there exists an unknown homography H between corresponding points for all $i \in \{1, 2, \dots, N\}$. This formally means that $\exists H$ such that:

$$(2)\ p^i \equiv Hq^i$$

where $p^i = (x_i, y_i, 1)$ and $q^i = (u_i, v_i, 1)$ are homogeneous coordinates of image points each from an image taken with C_1 and C_2 respectively.

- Given N correspondences in p and q and using Equation 2, derive a set of $2N$ independent linear equations in the form:

  $$(3)\ Ah = 0$$

  where h is a vector of the elements of H and A is a matrix composed of elements derived from the point coordinates.
  - Write down an expression for A.

Hint: Start by writing out Equation 2 in terms of the elements of H and the homogeneous coordinates for p^i and q^i.

- How many elements are there in h?
- How many point pairs (correspondences) are required to solve this system? Why?

Hint: How many degrees of freedom are in H? How much information does each point correspondence give?

Part 2.B - Planar Homographies: Practice

In this part you will implement an image stitching algorithm, you will learn how to stitch several images of the same scene into a panorama. First, we'll concentrate on the case of two images and then extend to several images.

For the following tasks (Part 2):

- **You are not allowed to use OpenCV/Scipy or any other "ready to use" functions when you are asked to implement a function (you can still use the functions to save and load images).**
- Add all your implemented functions for this section into `my_homography.py` under the `code` folder.
- Make sure you answer/demostrate all the "bullet" questions in your report
- For each step add to your reprot illustration images.
- You can demonstrate your steps using `/code/data/incline_L.jpg` and `/code/data/incline_R.jpg` images, or any other relevant example images (unless specified otherwise).

2.1 Manual finding corresponding points

Implement a function that enables manual matching of corresponding image points between two images. You will probably want to use the function `ginput()` from `matplotlib` . The success of the registration depends on the accuracy of your match, so mark it carefully. (It might be difficult to make `ginput()` work in a `Jupyter Notebook` . If you use `Pycharm` make sure your `Show plot in tool window` checkbox is off, you can find it under `Python Scientific` in the setting menu).

In [ ]:
```python
def getPoints(im1, im2, N):
    """
    Your code here
    """
    return p1,p2
```

Inputs: `im1` and `im2` are two 2D grayscale images. `N` is the number of corrosponding points you want to extract. Output: `p1` and `p2` should be $2 \times N$ matrices of corresponding $(x, y)^T$ coordinates between two images.

2.2 Calculate transformation:

Implement a function that gets a set of matching points between two images and calculates the transformation between them. The transformation should be 3×3 H homogenous matrix such that for each point in image $p \in C_1$, there would be a transformation in image C_2 such that $p = Hq$, $q \in C_2$.

In [ ]:
```python
def computeH(p1, p2):
    """
    Your code here
    """
    return H2to1
```

Inputs: `p1` and `p2` should be $2 \times N$ matrices of corresponding $(x, y)^T$ coordinates between two images. Outputs: `H2to1` should be a 3×3 matrix encoding the homography that best matches the linear equation derived above for Equation 2. Hint: Remember that a homography is only determined up to scale. The `numpy` 's functions `eig()` or `svd()` will be useful. Note that this function can be written without an explicit for-loop over the data points. *Hint for debugging*: A good test of your code is to check that the homography of an image with itself is an identity.

- Implement the H computation function. Describe and explain your implementation.
- Show that the transformation is correct by selecting arbitrary points in the first image and project them to the second image.

2.3 Image warping:

Implement a function that gets an input image and a transformation matrix `H` and returns the projected image. Please note that after the projection, there will be coordinates that won't be integers (e.g. sub-pixels). Therefore you will need to interpolate between neighboring pixels. For color images, project the image for each color channel and then connect them together. It is also better to use the `LAB` color space. In order to avoid holes, use inverse warping.

- Implement the wrapping function using `numpy` and `SciPy` `interp2d()` function .
- Discuss the influences of different interpolations kinds { `'linear'` , `'cubic'` }.
- Discuss the influence of using `LAB` color space.

Note: When performing algorithm with multiple steps, you need to demonstrate and explain each of those additianal improvments.

In [1]:
```python
def warpH(im1, H, out_size):
    """
    Your code here
    """
    return warp_im1
```

Inputs: `im1` is a colored image. `H` is a 3×3 matrix encoding the homography between im1 and im2. `out_size` is the size of the wanted output `(new_imH,new_imW)` . Output: `warp_im1` is the transposed warp image `im1` include empty background (zeros).

2.4 Panorama stitching:

Implement a function that gets two images after axis alignment and returns a union of the two. The union should be a simple overlay of one image on the other. Leave empty pixels painted black.

In [ ]:
```python
def imageStitching(img1, warp_img2):
    """
    Your code here
    """
    return panoImg
```

Inputs: `im1` , `warp_img2` are two colored images. Output: `panoImg` is the output gathered panorama.

- Use all the above functions to create a panorama image. Demonstrate and explain you results on the `/code/data/incline` images.

2.5 Autonomous panorama stitching using SIFT:

In this section, we are going to use SIFT to stitch images. Read https://docs.opencv.org/3.4/da/df5/tutorial_py_sift_intro.html (https://docs.opencv.org/3.4/da/df5/tutorial_py_sift_intro.html) to learn how to use the OpenCV to extract SIFT keypoints and descriptors. Now, you need to find corresponding points, matching points, you can use to following tutorial: https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_feature2d/py_matcher/py_matcher.html (https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_feature2d/py_matcher/py_matcher.html) Implement the `getPoints_SIFT()` function, which get two images and outputs `p1,p2` SIFT keypoints, where `p1[j],p2[j]` are pairs of cooresponding points between `im1` and `im2` .

In [ ]:
```python
def getPoints_SIFT(im1, im2):
    """
    Your code here
    """
    return p1,p2
```

- Choose a descriptors matching algorithm, and explain how it works (e.g. KNN).
- Use the SIFT and the match algorithm to perform image stitching over the `code/data/incline` images.

2.7 Compare SIFT and Manuale image selection:

- Show the results of the *manual* and *SIFT* matching algorithm on the attached images of the beach and Pena National Palace (Portugal) for the entire set of images.
  - Compare the two methods (pros/cons).

2.8 RANSAC:

The least squares method you implemented for computing homographies is not robust to outliers. If all the correspondences are good matches, this is not a problem. But even a single false correspondence can completely throw off the homography estimation. When correspondences are determined automatically (using BRIEF feature matching for instance), some mismatches in a set of point correspondences are almost certain. **RANSAC (Random Sample Consensus)** can be used to fit models robustly in the presence of outliers.

- Write a function that uses RANSAC to compute homographies automatically between two images:

In [ ]:
```
def ransacH(p1, p2, nIter, tol):
    """
    Your code here
    """
    return bestH
```

The inputs and output of this function should be as follows:

Inputs: `p1` and `p2` are matrices specifying point locations in each of the images and `p1[j],p2[j]` are matched points between two images.

Algorithm Input Parameters: `nIter` is the number of iterations to run RANSAC for, `tol` is the tolerance value for considering a point to be an inlier. Define your function so that these two parameters have reasonable default values.

Outputs: `bestH` should be the homography model with the most inliers found during RANSAC.

- Use the RANSAC to improve your image stitching results from the last section.
- Demonstrate the improvement and explain (for both manual and SIFT).

2.9 Blending - BONUS

Pay attention: this is a BONUS task

Improve your results visibility using blending (featering) the panorama's borders.

- Create a `blender()` function (you declare the input and outputs).
- Explain and demonstrate your blending function.

2.10 Be Creative:

- Go out (if you are allowed to...) and take at least 3 pictures of a far distance object (e.g. a building), use what you have learned to create a new excellent Panorama image.

2.11 Affine vs Projective - BONUS:

Pay attention: this is a BONUS task

- Create a function that compute Affine transformation. Describe yours steps in the PDF report.
- Find one example where we can use Affine transformation, and one where we can **not** use Affine transformation for panorama stitching, demonstrate (compare with Projective).

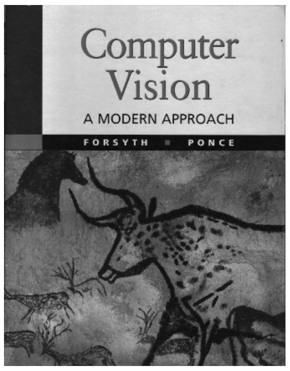
Part 3 - Creating your Augmented Reality application

Now with the code you have, you can create your own Augmented Reality application.

- Add all your implemented functions for this section into `my_ar.py` under the `code` folder.

3.1 Create reference model

Take a picture of your favourite book front cover (place it within `/code/my_data` folder). Make sure the book is not occluded, and the image resolution is reasonable. We want to convert your book to a model template. Use the functions you have created in the last section to project the front cover into the XY plane. Use your functions to get the books corners and project it to a reasenoble rectangle. The results should look like the following example:



```
In [ ]: def create_ref(im_path):
            """
               Your code here
            """
            return ref_image
```

`create_ref` gets the original image path and outputs a ref image.

- Add the resulted image to your report and to `code/my_data` folder.

3.2 Implant an image inside another image

Take another picture of the same book, this time plant the book somewhere in the scene. Now, take an image of another book (or something else) and stitch it within the first image, instead of the book, using planer homography.

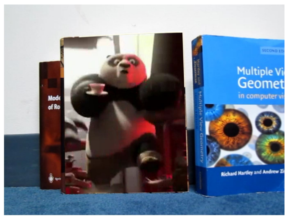
- Add your function to `my_ar.py` file as `im2im()` (declare yourself the input and outputs).
- Add at least 3 examples to your `output` folder as `im2imx.jpg` where `x` is the example number.
- Add one of the examples to your report.
- Explain all your steps.

3.3 Video time - BONUS

Pay attention: this is a BONUS task

Here you are going to create your Augmented Reality app. You are going to implant a video within another video.

Create a short video (less than 10 sec) of your favorite book, move the book or the camera during the video, and make sure the book is visible within all the frame. Now, choose another video that you like to implant inside the first one. Your result should be similar to the LifePrint (https://www.indiegogo.com/projects/lifeprint-photos-that-come-to-life-in-your-hands/#/) project. You'll be given full credits if you can put the video together correctly. See the following figure for an example frame of what the final video should look like.



Two notes: (a) The book and the videos you chose probably have very different aspect ratios (the ratio of the image width to the image height). You must either use `imresize` or `crop` each frame to fit onto the book cover. (b) You can replace the SIFT with ORB (https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_feature2d/py_orb/py_orb.html) for a faster implementation.

- Add your function to `my_ar.py` file as `my_vid2vid()` (declare yourself the input and outputs).
- Add your output video to the `output` folder as `vid2vid.mp4` .
- Add few illustration frames to your report (at least 3).
- Explain all your steps. Did you change anything from sec. `3.2` ? Why?

3.4 Creativity Section -MANDATORY (Not a Bonus)

If you have completed section 3.3, this is your task:

We encourage creativity, and we want you to add something creative to the final product. It can be anything from what we have taught you in the course. Here are some suggestions:

- Deep object detection/tracking. Can you detect/track objects in the final product? Can you track point and anchor a text to it?
- Style your video - use a pre-trained GAN (or any classical algorithm) to stylize your video (Pix2Pix and CycleGAN can be used).
- Segment an object from an image or video and add it to your video.

If you have not completed section 3.3, this is your task:

Extend your video from the previous HW (HW3) by adding something creative, **which is not another segmentation**. Here are some suggestions:

- Deep object detection/tracking. Can you detect/track objects in the final product? Can you track point and anchor a text to it?
- Style your video - use a pre-trained GAN (or any classical algorithm) to stylize your video (Pix2Pix and CycleGAN can be used).
- Use any algorithm learned in the course to add something creative (other than segmentation).

- In the report, describe the algorithms you used.
- Add your script to the `code` folder and name it `my_vid2vid_ext.py` (or `my_vid2vid_ext.ipynb`).
- Add your output video to the `output` folder as `vid2vid_ext.mp4` and upload it to YouTube as instructed above.
- Add few illustration frames to your report.

References & Credits

- Carnegie Mellon University - CMU
- Icons from Icon8.com (https://icons8.com/) - https://icons8.com (https://icons8.com)