

Problem Set 3: Introduction to AR

Description

Problem Set 3 introduces basic concepts behind Augmented Reality, using the contents that you will learn in modules 3A-3D and 4A-4C: Projective geometry, Corner detection, Perspective imaging, and Homographies, respectively.

Additionally, you will also learn how to read from a video, process each video frame by identifying important features, insert images within images, and assemble a video from a sequence of frames.

Learning Objectives

- Find markers using circle and corner detection, convolution, and / or pattern recognition.
- Learn how projective geometry can be used to transform a sample image from one plane to another.
- Address the marker recognition problem when there is noise in the scene.
- Implement backwards (reverse) warping.
- Understand how video can be extracted in sequences of images, and replace specific areas of each image with different content.
- Assemble a video from a sequence of images.

Problem Overview

Methods to be used: In this assignment you are to use methods that work with Feature Correspondence and Corner detection. You will also apply methods that are part of Projective Geometry and Image Warping, however you will have to do these manually using linear algebra concepts.

RULES: You may use image processing functions to find color channels, load images, find edges (such as with Canny). Don't forget that those have a variety of parameters and you may need to experiment with them. There are certain functions that may not be allowed and are specified in the assignment's autograder Piazza post. Refer to this problem set's autograder post for a list of banned function calls.

Please do not use absolute paths in your submission code. All paths should be relative to the submission directory. Any submissions with absolute paths are in danger of receiving a penalty!

Obtaining the Started Files

Obtain the starter code from canvas under files.

Programming instructions

Your main programming task is to complete the api described in the file **ps3.py**. The driver program **experiment.py** helps to illustrate the intended use and will output the files needed for the writeup. Additionally there is a file **ps3_test.py** that you can use to test your implementation.

Write-up instructions

Create **ps3_report.pdf** - a PDF file that shows all your output for the problem set, including images labeled appropriately (by filename, e.g. ps3-1-a-1.png) so it is clear which section they are for and the small number of written responses necessary to answer some of the questions (as indicated). For a guide as to how to showcase your results, please refer to the powerpoint template for PS3.

How to submit:

1. To submit your code, in the terminal window run the following command:
`python submit.py ps03`
2. To submit the report, input images for part 5, and experiment.py, in the terminal window run the following command:
`python submit.py ps03_report`
3. Submit your report pdf to gradescope.

YOU MUST PERFORM ALL THREE STEPS. I.e. two commands in the terminal window and one upload to gradescope. Only your last submission before the deadline will be counted for each of the code and the report.

The following lines will appear:

GT Login required.

Username : <GT username (same as T-square)>

Password: <GT password>

Save the jwt?[y,N] <either y or N if you want to save your credentials>

You should see the autograder's feedback in the terminal window. Additionally, you can look at a history of all your submissions at <https://bonnie.udacity.com/>

Grading

The assignment will be graded out of 100 points. The last submission before the time limit will only be considered. The code portion (autograder) represents **60%** of the grade and the report the remaining **40%**.

The images included in your report must be generated using experiment.py. This file should be set to be run as is to verify your results. **Your report grade will be affected if we cannot reproduce your output images.**

The report grade breakdown is shown in the question heading. As for the code grade, you will be able to see it in the console message you receive when submitting.

Assignment Overview

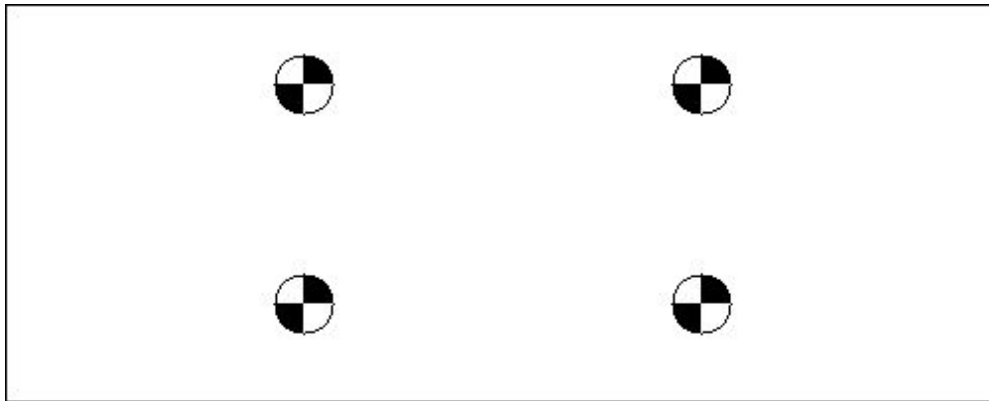
A glass/windshield manufacturer wants to develop an interactive screen that can be used in cars and eyeglasses. They have partnered with a billboard manufacturer in order to render certain marketing products according to each customer's preferences.

Their goal is to detect four points (markers) currently present in the screen's field-of-view and insert an image or video in the scene. To help with this task, the advertising company is installing blank billboards with four distinct markers, which determine the area's intended four corners. The advertising company plans to insert a target image / video into this space.

They have hired you to produce the necessary software to make this happen. They have set up their sensors so that you will receive an image / video feed and a target image / video. They expect an altered image / video that contains the target content rendered in the scene, visible in the screen.

1. Marker detection in a simulated scene [15 Points]

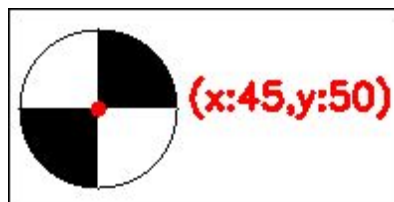
The first task is to identify the markers for this Augmented Reality exercise. In real practice, markers can be used (in the form of unique pictures) that stand out from the background of an image. Below is an image with four markers.



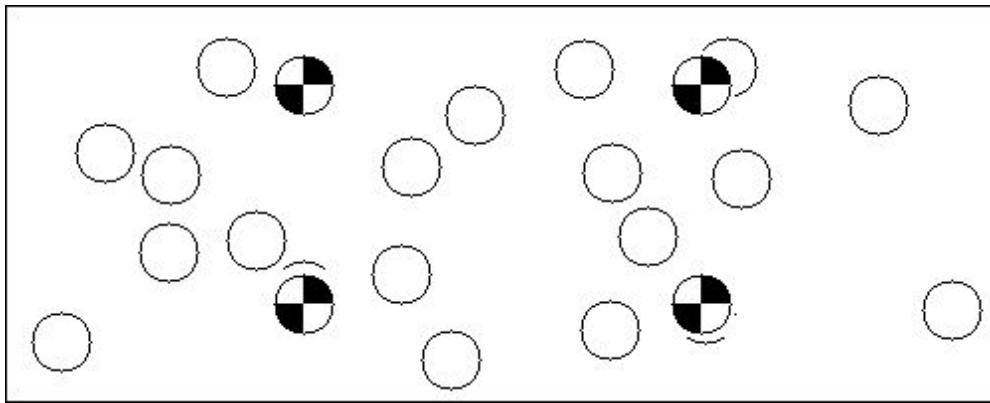
Notice that they contain a cross section bounded by a circle. The cross-section is useful in that it forms a distinguished corner. In this section you will create a function/set of functions that can detect these markers, as shown above. You will use the images provided to detect the (x, y) center coordinates of each of these markers in the image. The position should be represented by the center of the marker (where the cross-section is).

Code: Complete `find_markers(image)`

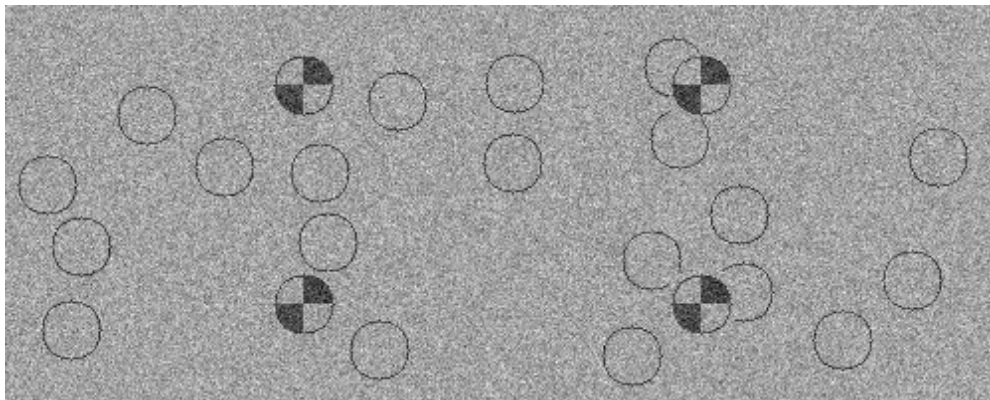
You will use the function `mark_location(image, pt)` in `experiment.py` to create a resulting image that highlights the center of each marker and overlays the marker coordinates in the image. Each marker should present their location similar to this:



Images like the one above may not be that hard to solve. However, in a real-life scene, it proves to be much more difficult. Make sure your methods are robust enough to also locate the markers in images like the one below, where there could be other objects in the scene:



Let's now assume there is "noise" in the scene (i.e. rain, fog, etc.).



Report: Find the markers and place their coordinates, as shown above. Use the following images:

- Input: **sim_clear_scene.jpg**. Output: **ps3-1-a-1.png**
- Input: **sim_noisy_scene_1.jpg**. Output: **ps3-1-a-2.png**
- Input: **sim_noisy_scene_2.jpg**. Output: **ps3-1-a-3.png**

2. Marker detection in a real scene [20 Points]

Now that you have a working method to detect markers in simulated scenes, you will adapt it to identify these same markers in real scenes like the image shown below. Use the images provided to essentially repeat the task of section 1 above and draw a box (four 3-pixel wide lines, any color) where the box corners touch the marker centers.



Code: Complete `draw_box(image, markers)`

Report: Find the markers and place their coordinates, as shown above. Use the following images:

- Input: **ps3-2-a_base.jpg**. Output: **ps3-2-a-1.png**
- Input: **ps3-2-b_base.jpg**. Output: **ps3-2-a-2.png**
- Input: **ps3-2-c_base.jpg**. Output: **ps3-2-a-3.png** (90-degree rotation is intentional)
- Input: **ps3-2-d_base.jpg**. Output: **ps3-2-a-4.png**
- Input: **ps3-2-e_base.jpg**. Output: **ps3-2-a-5.png**

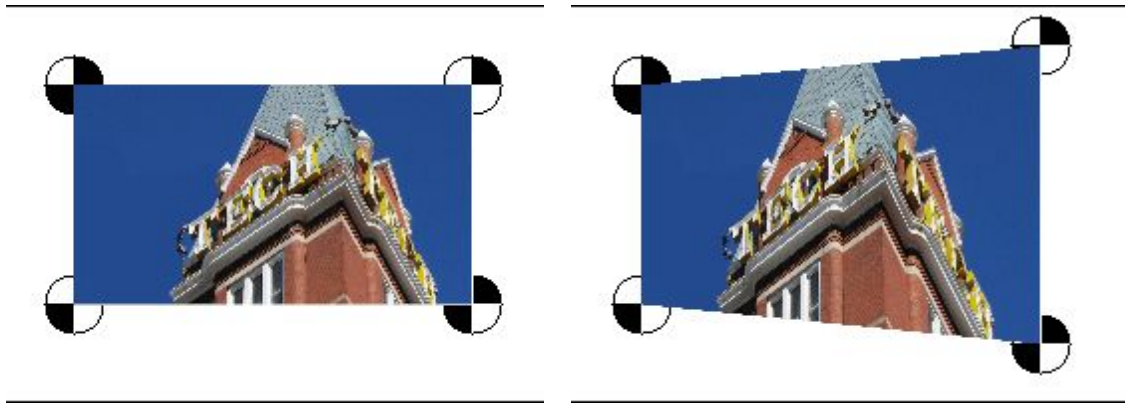
3. Projective Geometry [20 Points]

Now that you know where the billboard markers are located in the scene, we want to add the marketing image. The advertising company requires that their client's billboard image is visible from all possible angles since you are not just driving straight into the advertisements. Unphased, you know enough about computer vision to introduce projective geometry. The next task will use the information obtained in the previous section to compute a transformation matrix H . This matrix will allow you to project a set of points (x, y) to another plane represented by the points (x', y') in a 2D view. In other words we are looking at the following operation:

$$\begin{bmatrix} wx' \\ wy' \\ w \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

In this case, the 3x3 matrix is a *homography*, also known as a *perspective transform* or *projective transform*. There are eight unknowns, **a** through **h**, and **i** is 1. If we had four pairs of corresponding $(u, v) \leftrightarrow (u', v')$ points, we can solve for the homography.

The objective here is to insert an image in the rectangular area that the markers define. This insertion should be robust enough to support cases where the markers are not in an orthogonal plane from the point of view and present rotations. Here are two examples of what you should achieve:



Code: Complete:

- `find_four_point_transform(src_points, dst_points)`
- `project_imageA_onto_imageB(imageA, imageB, homography)`

Report: Use an image of your own to project in the area delimited by the four markers. Name it **img-3-a-1.png** and place it in the "input_images" directory.

- Input: **ps3-3-a_base.jpg, img-3-a-1.png**. Output: **ps3-3-a-1.png**
- Input: **ps3-3-b_base.jpg, img-3-a-1.png**. Output: **ps3-3-a-2.png**
- Input: **ps3-3-c_base.jpg, img-3-a-1.png**. Output: **ps3-3-a-3.png**

4. Finding markers in a video [20 Points]

Static images are fine in theory, but the company wants this functional and put into practice. That means, finding markers in a moving scene.

In this part you will work with a short video sequence of a similar scene. When processing videos, you will read the input file and obtain images (frames). Once the image is obtained, you will apply the same concept as explained in the previous sections. Unlike the static image, the input video will change in translation, rotation, and perspective. Additionally there may be cases where a few markers are partially visible. Finally, you will assemble this collection of modified images into a new video. Your output must render each marker position relative to the current frame coordinates.

Besides making all the necessary modifications to make your code more robust, you will complete a function that outputs a video frame generator. This function is almost complete and it is placed so that you can learn how videos are read using OpenCV. Follow the instructions placed in `ps3.py`.

Code: Complete `video_frame_generator(filename)`

Report: In order to grade your implementation, share a link to each video (Youtube, Dropbox, etc.). This video should be only visible via link sharing, not public. If we cannot open this link when grading your grade for this and remaining sections will be affected.

- a. First we will start with the following videos. Include the specified frames in your report as instructed below.

Frames to record: 355, 555, and 725.

- Input: **ps3-4-a.mp4**. Output: **ps3-4-a-1.png, ps3-4-a-2.png, ps3-4-a-3.png, link to the full video.**

Frames to record: 97, 407, and 435.

- Input: **ps3-4-b.mp4**. Output: **ps3-4-a-4.png, ps3-4-a-5.png, ps3-4-a-6.png, link to the full video.**

- b. Now work with noisy videos:

Frames to record: 47, 470, and 691.

- Input: **ps3-4-c.mp4**. Output: **ps3-4-b-1.png, ps3-4-b-2.png, ps3-4-b-3.png, link to the full video.**

Frames to record: 207, 367, and 737.

- Input: **ps3-4-d.mp4**. Output: **ps3-4-b-4.png, ps3-4-b-5.png, ps3-4-b-6.png, link to the full video.**

5. Final Augmented Reality [20 Points]

Now that you have all the pieces, insert your advertisement into the video provided. Pick an image and insert it in the provided video.

Report: In order to grade your implementation, share a link to each video (Youtube, Dropbox, etc.). This video should be only visible via link sharing, not public. If we cannot open this link when grading your grade for this and remaining sections will be affected.

- c. First we will start with the following videos. Include the specified frames in your report as instructed below.

Frames to record: 355, 555, and 725.

- Input: **ps3-4-a.mp4**. Output: **ps3-5-a-1.png, ps3-5-a-2.png, ps3-5-a-3.png, link to the full video.**

Frames to record: 97, 407, and 435.

- Input: **ps3-4-b.mp4**. Output: **ps3-5-a-4.png, ps3-5-a-5.png, ps3-5-a-6.png, link to the full video.**

- d. Now work with noisy videos:

Frames to record: 47, 470, and 691.

- Input: **ps3-4-c.mp4**. Output: **ps3-5-b-1.png, ps3-5-b-2.png, ps3-5-b-3.png, link to the full video.**

Frames to record: 207, 367, and 737

- Input: **ps3-4-d.mp4**. Output: **ps3-5-b-4.png, ps3-5-b-5.png, ps3-5-b-6.png, link to the full video.**

6. Challenge problem: Video in Video [5 points]

As a challenge, try embedding a video inside the markers video. You are free to select any video and modify it as necessary to make it fit both in size and number of frames. Name this video **my-ad.mp4**, this file will not be collected as it may exceed the Bonnie submission size limit. A different file with the same name will be used when grading your assignment (which shouldn't affect your results). The file we will use for grading is longer in

duration than ps3-4-a.mp4. Your output should have the same size and number of frames as the original markers video.

- Frames to record: 355, 555, and 725.
- Input: **ps3-4-a.mp4, my-ad.mp4**. Output: **ps3-6-a-1.png, ps3-6-a-2.png, ps3-6-a-3.png, link to the full video.**

