COMP4901L Computer Vision, Fall 2019

Homework Assignment 2
Augmented Reality with Planar Homographies

In this assignment, you will be implementing an AR application step by step using planar homographies. Before we step into the implementation, we will walk you through the theory of planar homographies. In the programming section, you will first learn to find point correspondences between two images and use these to estimate the homography between them. Using this homography you will then warp images and finally implement your own AR application.

Instructions

1. **Integrity and collaboration:** Students are encouraged to work in groups but each student must submit their own work. If you work as a group, include the names of your collaborators in your write up. Code should NOT be shared or copied. Please DO NOT use external code unless permitted. Plagiarism is strongly prohibited and may lead to failure of this course.

2. **Start early!** Especially those not familiar with Matlab.

3. **Write-up:** Your write-up should mainly consist of three parts, your answers to theory questions, resulting images of each step, and the discussions for experiments. Please note that we DO NOT accept handwritten scans for your write-up in this assignment. Please type your answers to theory questions and discussions for experiments electronically.

4. Please stick to the function prototypes mentioned in the handout. This makes verifying code easier for the TAs.

5. **Submission:** Create a zip file, `<ust-login-id>.zip`, composed of your write-up, your Matlab implementations (including helper functions), and your implementations, results for extra credits (optional). Please make sure to remove the data `/folder`, `loadVid.m`, `warpH.m` and any other temporary files you've generated. Your final upload should have the files arranged in this layout:

   `<ust_login_id>.zip`

   - `<ust-login-id>/`
     - `<ust-login-id>.pdf`
     - `matlab/`
       - `MatchPics.m`
       - `briefRotTest.m`
       - `computeH.m`
       - `computeH norm.m`
       - `computeH ransac.m`
       - `compositeH.m`
       - `HarryPotterize_auto.m`
       - `ar.m`
       - `yourHelperFunctions.m` *(optional)*
     - `result/`

Figure 1: A homography $\mathbf{H}$ links all points $\mathbf{x}_\pi$ lying in plane π between two camera views $\mathbf{x}$ and $\mathbf{x}'$ in cameras C and C' respectively such that $\mathbf{x}' = \mathbf{H}\mathbf{x}$. [From Hartley and Zisserman]

    - `ar.avi`
  - `ec/` (optional for extra credit)
    - `ar_ec.m`
    - `panorama.m`
    - the images required for generating the results.

Please make sure you do follow the submission rules mentioned above before uploading your zip file to CASS. Assignments that violate this submission rule will be **penalized by up to 10% of the total score**.

6. **File paths:** Please make sure that any file paths that you use are relative and not absolute. Not `imread('/name/Documents/subdirectory/hw1/data/xyz.jpg')` but `imread('../data/xyz.jpg')`.

1 Homographies

Planar Homographies as a Warp

Recall that a planar homography is an warp operation (which is a mapping from pixel coordinates from one camera frame to another) that makes a fundamental assumption of the points lying on a plane in the real world. Under this particular assumption, pixel coordinates in one view of the points on the plane can be *directly* mapped to pixel coordinates in another camera view of the same points.

Q1.1 Homography (5 points)

Prove that there exists a homography $\mathbf{H}$ that satisfies equation 1 given two 3×4 camera projection matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ corresponding to the two cameras and a plane $\prod$. You do not need to produce an actual algebraic expression for $\mathbf{H}$. All we are asking for is a proof of the existence of $\mathbf{H}$.

$$\mathbf{x}_1 \equiv \mathbf{H}\mathbf{x}_2 \tag{1}$$

The $\equiv$ symbol stands for identical to. The points $\mathbf{x}_1$ and $\mathbf{x}_2$ are in homogenous coordinates, which means they have an additional dimension. If $\mathbf{x}_1$ is a 3D vector $[x_i \; y_i \; z_i]^T$, it represents the 2D point $[\frac{x_i}{z_i} \; \frac{y_i}{z_i}]^T$ (called inhomogenous coordinates). This additional dimension is a mathematical

convenience to represent transformations (like translation, rotation, scaling, etc) in a concise matrix form. The $\equiv$ means that the equation is correct to a scaling factor.

Note: A degenerate case happens when the plane $\prod$ contains both cameras' centers, in which case there are infinite choices of H satisfying equation 1. You can ignore this special case in your answer.

The Direct Linear Transform

A very common problem in projective geometry is often of the form $x \equiv \mathbf{A}\mathbf{y}$, where x and y are known vectors, and $\mathbf{A}$ is a matrix which contains unknowns to be solved. Given matching points in two images, our homography relationship clearly is an instance of such a problem. Note that the equality holds only *up to scale* (which means that the set of equations are of the form $x = \lambda \mathbf{H}\mathbf{x}'$), which is why we cannot use an ordinary least squares solution such as what you may have used in the past to solve simultaneous equations. A standard approach to solve these kinds of problems is called the Direct Linear Transform, where we rewrite the equation as proper homogeneous equations which are then solved in the standard least squares sense. Since this process involves disentangling the structure of the $\mathbf{H}$ matrix, it's a *transform* of the problem into a set of *linear* equations, thus giving it its name.

Q1.2 Correspondences (15 points)

Let $\mathbf{x}_1$ be a set of points in an image and $\mathbf{x}_2$ be the set of corresponding points in an image taken by another camera. Suppose there exists a homography $\mathbf{H}$ such that:

$$\mathbf{x}_1^i \equiv \mathbf{H}\mathbf{x}_2^i \qquad (i \in \{1 \cdots N\})$$

where $\mathbf{x}_1^i = [x_1^i \ y_1^i 1]^T$ are in homogenous coordinates, $\mathbf{x}_1^i \in \mathbf{x}_1$ and $\mathbf{H}$ is a 3×3 matrix. For each point pair, this relation can be rewritten as

$$\mathbf{A}_i \mathbf{h} = 0$$

where $\mathbf{h}$ is a column vector reshaped from $\mathbf{H}$, and $\mathbf{A}_i$ is a matrix with elements derived from the points $\mathbf{x}_1^i$ and $\mathbf{x}_2^i$. This can help calculate $\mathbf{H}$ from the given point correspondences.

1. How many degrees of freedom does $\mathbf{h}$ have? (3 points)
2. How many point pairs are required to solve $\mathbf{h}$? (2 points)
3. Derive $\mathbf{A}_i$. (5 points)
4. When solving $\mathbf{A}\mathbf{h} = 0$, in essence you're trying to find the $\mathbf{h}$ that exists in the null space of $\mathbf{A}$. What that means is that there would be some non-trivial solution for $\mathbf{h}$ such that that product $\mathbf{A}\mathbf{h}$ turns out to be 0.
   What will be a trivial solution for $\mathbf{h}$? Is the matrix $\mathbf{A}$ full rank? Why/Why not?
   What impact will it have on the eigen values? What impact will it have on the eigen vectors? (5 points)

Using Matrix Decompositions to calculate the homography

A homography $\mathbf{H}$ transforms one set of points (in homogenous coordinates) to another set of points. In this project, we will obtain the corresponding point coordinates using feature matches and will then need to calculate the homography. You have already derived that $\mathbf{A}\mathbf{x} = \mathbf{0}$ in Question 1. In this section, we will look at how to solve such equations using two approaches, either of which can be used in the subsequent assignment questions.

Eigenvalue Decomposition

One way to solve $\mathbf{Ax} = \mathbf{0}$ is to calculate the eigenvalues and eigenvectors of $\mathbf{A}$. The eigenvector corresponding to $\mathbf{0}$ is the answer for this. Consider this example:

$$\mathbf{A} = \begin{bmatrix} 3 & 6 & -8 \\ 0 & 0 & 6 \\ 0 & 0 & 2 \end{bmatrix}$$

Using the Matlab function `eig`, we get the following eigenvalues and eigenvectors:

$$\begin{aligned} \mathbf{V} &= \begin{bmatrix} 1.0000 & -0.8944 & -0.9535 \\ 0 & 0.4472 & 0.2860 \\ 0 & 0 & 0.0953 \end{bmatrix} \\ \mathbf{D} &= \begin{bmatrix} 3 & 0 & 2 \end{bmatrix} \end{aligned}$$

Here, the columns of $\mathbf{V}$ are the eigenvectors and each corresponding element in $\mathbf{D}$ its eigenvalue. We notice that there is an eigenvalue of 0. The eigenvector corresponding to this is the solution for the equation $\mathbf{Ax} = \mathbf{0}$.

$$\mathbf{Ax} = \begin{bmatrix} 3 & 6 & -8 \\ 0 & 0 & 6 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} -0.8944 \\ 0.4472 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Singular Value Decomposition

The Singular Value Decomposition (SVD) of a matrix A is expressed as:

$$\mathbf{A} = U \Sigma T^T$$

Here, U is a matrix of column vectors called the "left singular vectors". Similarly, V is called the "right singular vectors". The matrix Σ is a diagonal matrix. Each diagonal element σ_i is called the "singular value" and these are sorted in order of magnitude. In our case, it is a 9×9 matrix.

- If $\sigma_9 = 0$, the system is *exactly-determined*, a homography exists and all points fit exactly.
- If $\sigma_9 \geq 0$, the system is *over-determined*. A homography exists but not all points fit exactly (they fit in the least-squares error sense). This value represents the goodness of fit.
- Usually, you will have at least four correspondences. If not, the system is *underdetermined*. We will not deal with those here.

The columns of U are eigenvectors of $\mathbf{AA}^T$. The columns of V are the eigenvectors of $\mathbf{A}^T\mathbf{A}$. We can use this fact to solve for $\mathbf{h}$ in the equation $\mathbf{Ah} = \mathbf{0}$. Using this knowledge, let us reformulate our problem of solving $\mathbf{Ax} = \mathbf{0}$. We want to minimize the error in solution in the least-squares sense. Ideally, the product $\mathbf{Ah}$ should be 0. Thus the sum-squared error can be written as:

$$\begin{aligned} f(\mathbf{h}) &= \frac{1}{2}(\mathbf{Ah} - \mathbf{0})^T(\mathbf{Ah} - \mathbf{0}) \\ &= \frac{1}{2}(\mathbf{Ah})^T(\mathbf{Ah}) \\ &= \frac{1}{2}\mathbf{h}^T\mathbf{A}^T\mathbf{Ah} \end{aligned}$$

Minimizing this error with respect to $\mathbf{h}$, we get:

$$\begin{aligned} \frac{d}{d\mathbf{h}} f &= 0 \\ \frac{1}{2}(\mathbf{A}^T\mathbf{A} + (\mathbf{A}^T\mathbf{A})^T)\mathbf{h} &= 0 \\ \mathbf{A}^T\mathbf{A}\mathbf{h} &= 0 \end{aligned}$$

This implies that the value of $\mathbf{h}$ equals the eigenvector corresponding to the zero eigenvalue (or closest to zero in case of noise). Thus, we choose the smallest eigenvalue of $\mathbf{A}^T\mathbf{A}$, which is σ_9 in Σ and the least-squares solution to $\mathbf{A}\mathbf{h} = \mathbf{0}$ is the the corresponding eigenvector (in column 9 of the matrix $\mathbf{V}$).

Theory Questions

Q1.3 Homography under rotation (5 points)

Prove that there exists a homography $\mathbf{H}$ that satisfies $\mathbf{x}_1 \equiv \mathbf{H}\mathbf{x}_2$, given two cameras separated by a pure rotation. That is, for camera 1, $\mathbf{x}_1 = \mathbf{K}_1[\mathbf{I}\ \mathbf{0}]\mathbf{X}$ and for camera 2, $\mathbf{x}_2 = \mathbf{K}_2[\mathbf{R}\ \mathbf{0}]\mathbf{X}$. Note that $\mathbf{K}_1$ and $\mathbf{K}_2$ are the 3×3 intrinsic matrices of the two cameras and are different. $\mathbf{I}$ is 3×3 identity matrix, $\mathbf{0}$ is a 3×1 zero vector and $\mathbf{X}$ is a point in 3D space. $\mathbf{R}$ is the 3×3 rotation matrix of the camera.

Q1.4 Understanding homographies under rotation (5 points)

Suppose that a camera is rotating about its center $\mathbf{C}$, keeping the intrinsic parameters $\mathbf{K}$ constant. Let $\mathbf{H}$ be the homography that maps the view from one camera orientation to the view at a second orientation. Let θ be the angle of rotation between the two. Show that $\mathbf{H}^2$ is the homography corresponding to a rotation of 2θ. Please limit your answer within a couple of lines. A lengthy proof indicates that you're doing something too complicated (or wrong).

Q1.5 Limitations of the planar homography (5 points)

Why is the planar homography not completely sufficient to map any arbitrary scene image to another viewpoint? State your answer concisely in one or two sentences.

Q1.6 Behavior of lines under perspective projections (5 points)

We stated in class that perspective projection preserves lines (a line in 3D is projected to a line in 2D). Verify algebraically that this is the case, i.e., verify that the projection $\mathbf{P}$ in $\mathbf{x} = \mathbf{P}\mathbf{X}$ preserves lines.

2 Computing Planar Homographies

Feature Detection and Matching

Before finding the homography between an image pair, we need to find corresponding point pairs between two images. But how do we get these points? One way is to select them manually (using `cpselect`), which is tedious and inefficient. The CV way is to find interest points in the image pair and automatically match them. In the interest of being able to do cool stuff, we will not reimplement a feature detector or descriptor here, but use built-in MATLAB methods. The purpose of an interest point detector (e.g. Harris, SIFT, SURF, etc.) is to find particular salient points in the images around which we extract feature descriptors (e.g. MOPS, etc.). These descriptors try to summarize the content of the image around the feature points in as succint yet descriptive manner

possible (there is often a tradeoff between representational and computational complexity for many computer vision tasks; you can have a very high dimensional feature descriptor that would ensure that you get good matches, but computing it could be prohibitively expensive). Matching, then, is a task of trying to find a descriptor in the list of descriptors obtained after computing them on a new image that best matches the current descriptor. This could be something as simple as the Euclidean distance between the two descriptors, or something more complicated, depending on how the descriptor is composed. For the purpose of this exercise, we shall use the widely used FAST detector in concert with the BRIEF descriptor.

Figure 2: A few matched FAST feature points with the BRIEF descriptor.

Q2.1.1 FAST Detector (5 points)

How is the FAST detector different from the Harris corner detector that you've seen in the lectures? (You will probably need to look up the FAST detector online.) Can you comment on its computational performance vis-à-vis the Harris corner detector?

Q2.1.2 BRIEF Descriptor (5 points)

How is the BRIEF descriptor different from the filterbanks you've seen in the lectures? Could you use any one of those filter banks as a descriptor?

Q2.1.3 Matching Methods (5 points)

The BRIEF descriptor belongs to a category called binary descriptors. In such descriptors the image region corresponding to the detected feature point is represented as a binary string of 1s and 0s. A commonly used metric used for such descriptors is called the *Hamming distance*. Please search online to learn about Hamming distance and *Nearest Neighbor*, and describe how they can be used to match interest points with BRIEF descriptors. What benefits does the Hamming distance distance have over a more conventional Euclidean distance measure in our setting?

Q2.1.4 Feature Matching (10 points)

Please implement a function:

```
[locs1, locs2] = matchPics(I1, I2)
```

where `I1` and `I2` are the images you want to match. `locs1` and `locs2` are $N \times 2$ matrices containing the x and y coordinates of the matched point pairs. Use the provided function `fast_corner_detect_9(im, threshold)` to compute the features (using a higher threshold gives you less points), then build descriptors using the provided `computeBrief` function and finally compare them using the provided method `customMatchFeatures`.

Use the function `customShowMatchedFeatures(im1, im2, locs1, locs2)` to visualize your matched

points and include the result image in your write-up. An example is shown in Figure 2. We provide you with the function:

```
[desc, locs] = computeBrief(img, locs in)
```

which computes the BRIEF descriptor for `img`. `locs_in` is an $N \times 2$ matrix in which each row represents the location (x, y) of a feature point. Please note that the number of valid output feature points could be less than the number of input feature points. `desc` is the corresponding matrix of BRIEF descriptors for the interest points.

Q2.1.5 BRIEF and Rotations (10 points)

Let's investigate how BRIEF works with rotations. Write a script `briefRotTest.m` that:

- Takes the `cv_cover.jpg` and matches it to itself rotated [Hint: use `imrotate`] in increments of 10 degrees.
- Stores a histogram of the count of matches for each orientation.
- Plots the histogram using `plot`.

Visualize the feature matching result at three different orientations and include them in your write-up. Explain why you think the BRIEF descriptor behaves this way. What happens when you switch to a different feature detector? [Hint: For instance, SURF features]. Would the plot change significantly? Why/Why not?

Homography Computation

Q2.2.1 Computing the Homography (15 points)

Write a function computeH that estimates the planar homography from a set of matched point pairs.

```
function [H2to1] = computeH(x1, x2)
```

`x1` and `x2` are $N \times 2$ matrices containing the coordinates (x, y) of point pairs between the two images. `H2to1` should be a 3×3 matrix for the best homography from image 2 to image 1 in the least-square sense. You can use `eig` or `svd` to get the eigenvectors (see Section 1 of this handout for details).

Homography Normalization

Normalization improves numerical stability of the solution and you should always normalize your coordinate data. Normalization has two steps:

1. Translate the mean of the points to the origin.
2. Scale the points so that the largest distance to the origin is $\sqrt{2}$.

This is a linear transformation and can be written as follows:

$$\begin{aligned} \widetilde{\mathbf{x}_1} &= \mathbf{T}_1 \mathbf{x}_1 \\ \widetilde{\mathbf{x}_2} &= \mathbf{T}_2 \mathbf{x}_2 \end{aligned}$$

where $\widetilde{\mathbf{x}_1}$ and $\widetilde{\mathbf{x}_2}$ are the normalized homogeneous coordinates of $\mathbf{x}_1$ and $\mathbf{x}_2$. $\mathbf{T}_1$ and $\mathbf{T}_2$ are 3×3 matrices. The homography H from $\widetilde{\mathbf{x}_2}$ and $\widetilde{\mathbf{x}_1}$ computed by computeH satisfies:

$$\widetilde{\mathbf{x}_1} = \mathbf{H}\widetilde{\mathbf{x}_2}$$

By substituting $\widetilde{\mathbf{x}_1}$ and $\widetilde{\mathbf{x}_2}$ with $\mathbf{T}_1\mathbf{x}_1$ and $\mathbf{T}_2\mathbf{x}_2$, we have:

$$\begin{aligned} \mathbf{T}_1\widetilde{\mathbf{x}_1} &= \mathbf{H}\mathbf{T}_2\widetilde{\mathbf{x}_2} \\ \mathbf{x}_1 &= \mathbf{T}_1^{-1}\mathbf{H}\mathbf{T}_2\widetilde{\mathbf{x}_2} \end{aligned}$$

Q2.2.2 Homography with normalization (10 points)

Implement the function `computeH norm`:

```
function [H2to1] = computeH norm(x1, x2).
```

This function should normalize the coordinates in `x1` and `x2` and call `computeH(x1, x2)` as described above.

RANSAC

The RANSAC algorithm can generally fit any model to noisy data. You will implement it for (planar) homographies between images. Remember that 4 point-pairs are required at a minimum to compute a homography.

Q2.2.3 Implement RANSAC for computing a homography (25 points)

Write a function:

```
function [bestH2to1, inliers] = computeH ransac(locs1, locs2)
```

where `bestH2to1` should be the homography $\mathbf{H}$ with most inliers found during RANSAC. $\mathbf{H}$ will be a homography such that if $\mathbf{x}_2$ is a point in `locs2` and $\mathbf{x}_1$ is a corresponding point in `locs1`, then $\mathbf{x}_1 \equiv \mathbf{H}\mathbf{x}_2$. `locs1` and `locs2` are $N \times 2$ matrices containing the matched points. `inliers` is a vector of length N with a 1 at those matches that are part of the consensus set, and 0 elsewhere. Use `computeH_norm` to compute the homography.

Automated Homography Estimation and Warping

Q2.2.4 Putting it together (10 points)

Write a script `HarryPotterize.m` that

1. Reads `cv_cover.jpg`, `cv_desk.png`, and `hp_cover.jpg`.
2. Computes a homography automatically using `MatchPics` and `computeH_ransac`.
3. Warps `hp_cover.jpg` to the dimensions of the `cv_desk.png` image using the provided `warpH` function.
4. At this point you should notice that although the image is being warped to the correct location, it is not filling up the same space as the book. Why do you think this is happening? How would you modify `hp_cover.jpg` to fix this issue?
5. Implement the function:

Figure 3: Text book

Figure 4: HarryPotterized Text book

```
function [ composite img ] = compositeH( H2to1, template, img )
```

to now compose this warped image with the desk image as in Figure 4.

6. Include your result in your write-up.

3 Creating your Augmented Reality application

Q3.1 Incorporating video (20 points)

Now with the code you have, you're able to create you own Augmented Reality application. What you're going to do is HarryPoterize the video `ar_source.mov` onto the video `book.mov`. More specifically, you're going to track the computer vision text book in each frame of `book.mov`, and overlay each frame of `ar_source.mov` onto the book in `book.mov`. Please write a script `ar.m` to implement this AR application and save your result video as `ar.avi` in the `result/` directory. You may use the function `loadVid.m` that we provide to load the videos. Your result should be similar to the LifePrint project. You'll be given full credits if you can put the video together correctly. See Figure 5 for an example frame of what the final video should look like.

Note that the book and the videos we have provided 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.

Cropping an image in Matlab is easy. You just need to extract the rows and columns you are interested in. For example, if you want to extract the subimage from point $(40, 50)$ to point $(100, 200)$, your code would look like `img_cropped = img(50:200, 40:100)`. In this project, you must crop that image such that only the central region of the image is used in the final output. See Figure 6 for an example.

Also, the video `book.mov` only has translation of objects. If you want to account for rotation of objects, scaling, etc, you would have to pick a better feature point representation (like ORB).

Figure 5: Rendering video on a moving target

Figure 6: Crop out the yellow regions of each frame to match the aspect ratio of the book

4 Extra Credit

Q4.1x: Make Your AR Real Time (15 points)

Write a script `ar_ec.m` that implements the AR program described in Q3.1 in real time. As an output of the script, you should process the videos frame by frame and have the combined frames played in real time. You don't need to save the result video for this question. The extra credits will be given to fast programs measured by FPS (frames per second). More specifically, we give 5 points to programs that run faster than 3 FPS, 10 points to programs running faster than 5 FPS and 15 points to programs running faster than 7 FPS.

Include the FPS and describe your techniques used for speeding up in your write-up. You should record the FPS **in the virtual barn environment**, computed as the total processing time divided by the total number of frames. You may exclude the data loading and displaying time.

Q4.2x: Create a Simple Panorama (10 points)

Take two pictures with your own camera, separated by a pure rotation as best as possible, and then construct a panorama with `panorama.m`. Be sure that objects in the images are far enough away so that there are no parallax effects. You can use Matlab's `cpselect` to select matching points on each image or some automatic method. Submit the original images, the final panorama, and the script `panorama.m` that loads the images and assembles a panorama. We have provided

two images for you to get started (`data/pano_left.png` and `data/pano_right.png`). Please use your own images when submitting this project. In your submission, include your original images and the panorama result image in your write-up. See the figures below for example.

Figure 7: Original Image 1 (left)

Figure 8: Original Image 2 (right)

Figure 9: Panorama