Homework 1

CSCI 497P / 597P, Winter 2019

1. 1D Filters

A 1D mean filter of width 3 (appropriate for applying to a 1D image) looks like:

1/3	1/3	1/3

Define a single 1D filter that, when applied only once to an image, will produce the same results as applying the 1D width-3 mean filter above twice. Write your filter in the space below

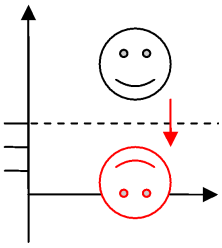
Kernel

2. Suppose you have a 64x64 image and you wish to build a 4-level Gaussian pyramid with a scale factor of 2 using a 3x3 kernel. Assume you're using 'same' mode and reflected borders for convolution padding. How many multiplications between image and kernel values are performed in total by all the convolution operations together?

3. Specify the homography (a 3x3 matrix) corresponding to each kind of warp:

a) A uniform scale by s plus translation by (u, v): $x' = sx + u$; $y' = sy + v$

b) A mirror reflection about the line y = 3, as shown at right. It may be helpful if you think of this as a sequence (multiplication) of simpler operations. Show your work.



4. Consider the following three matrices, where the elements a-h are non-zero:

$$A = \begin{bmatrix} a & b & 0 \\ d & e & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$B = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} a & b & c \\ d & e & f \\ h & g & 1 \end{bmatrix}$$

For each matrix, specify which matrices can perform each of the following transformations:

a. Translation:

b. Rotation about the origin:

c. Rotation about an arbitrary point:

d. Correct perspective distortion:

e. Alter the angle between two lines:

f. Shear:

g. Scale:

h. Map parallel lines to non-parallel lines:

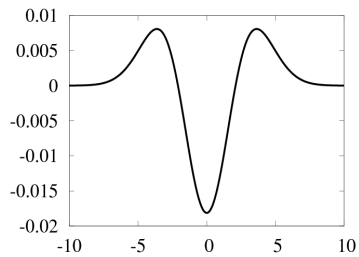
i. Map the origin to something other than (0,0):

j. Change the lengths of lines:

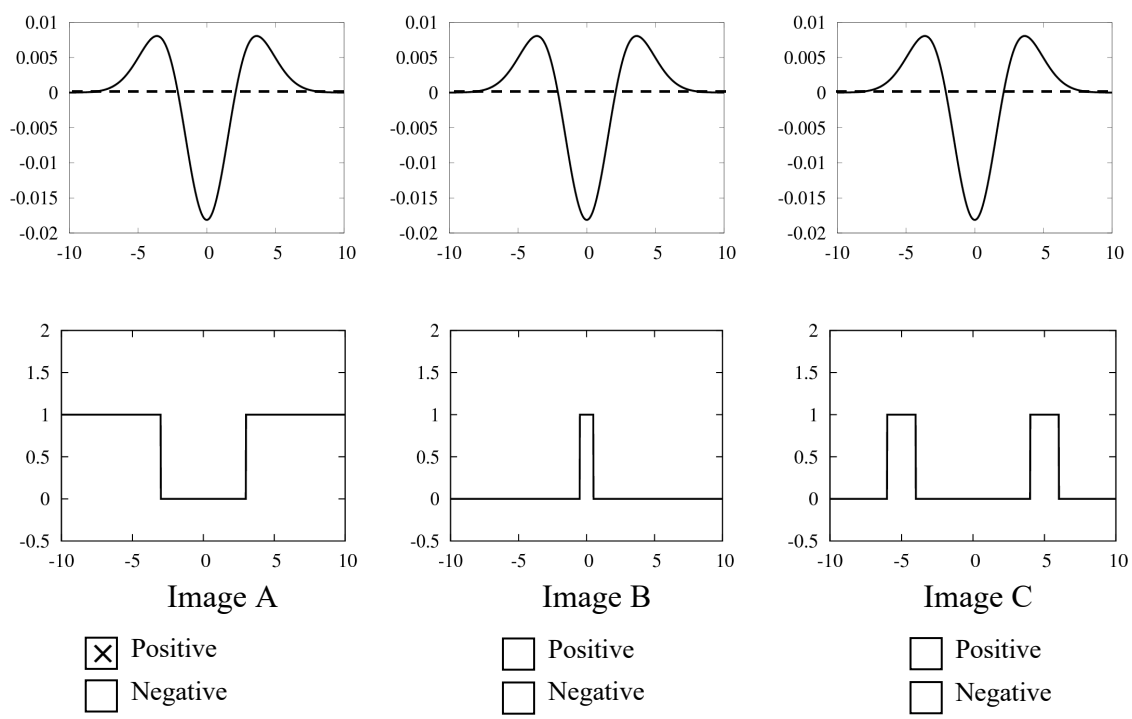
k. Change the ratios of lengths of lines:

5. SIFT features

While the Harris corner detector finds features which are maxima of the Harris operator, the SIFT detector finds features that are maxima (or minima) of a different function. SIFT first convolves the images with a *difference of Gaussians* (DoG) filter, which is similar to the Laplacian of Gaussian filter described in class. SIFT then finds maxima and minima of the convolved image. In 1D, the DoG filter looks like this (note that some values are negative):



Consider the three images below labeled Image A, Image B, and Image C. This question asks you to apply the 1D DoG filter to each image at the position x=0, and determine the sign of the resulting number (called the *response* of the filter). For which images is the filter response at x=0 positive? For which images is it negative? Check one of the boxes for each image; the correct box for Image A has already been checked for you. (The DoG filter has been replicated above each image to help you visualize the convolution). Note that the x-axis of each 1D image represents position, and the y-axis represents intensity (an intensity of 0 indicates black, and an intensity of 1 indicates white).



6. In the Harris detector, why is it necessary to use eigen-analysis of the H matrix computed from a window around a pixel to compute corner strength, rather than simply looking at the gradient magnitudes in the same window?

7. Give two examples of transformations that the Harris corner detector is *not* invariant to.

8. In Project 2, you implemented functionality for matching local features between two images (call them A and B). In particular, to match features in A to B, for each feature in image A, you found the nearest neighbor feature (in descriptor space) in image B. Would you find the same matches if you matched features in B to A (i.e., reversed the role of the two images)? Why or why not?

9. In class we studied RANSAC as an approach for robustly estimating transforms between images given noisy feature matches contaminated with outliers. One RANSAC parameter is number of iterations (i.e., the number of hypotheses generated) before we terminate and accept the best transform found so far (which is hopefully also the correct transform).

Recall that we derived the following formula for calculating the number of iterations **K** necessary to successfully fit a model with probability **P**, given that the inlier ratio is **r** and the minimum number of samples to fit a model is **s**:

$$K \leq \frac{\ln(1-P)}{\ln(1-r^s)}$$

You are tasked with developing a real-time mobile video stabilization algorithm that must fit a motion model to every pair of frames on a limited computational budget. In particular, you have time to perform only **K** = 20 RANSAC iterations. Assuming the inlier ratio (fraction of matches that are inliers) **r** is 0.5, give the probability of fitting each of the following transformations correctly:

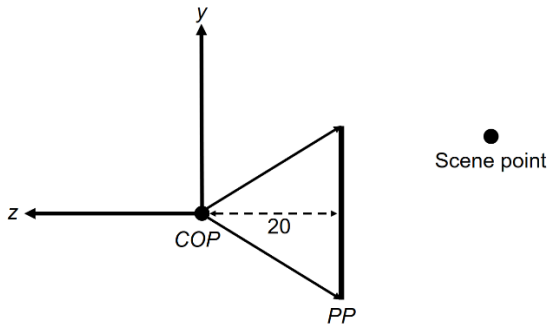
a. Translation

b. Affine

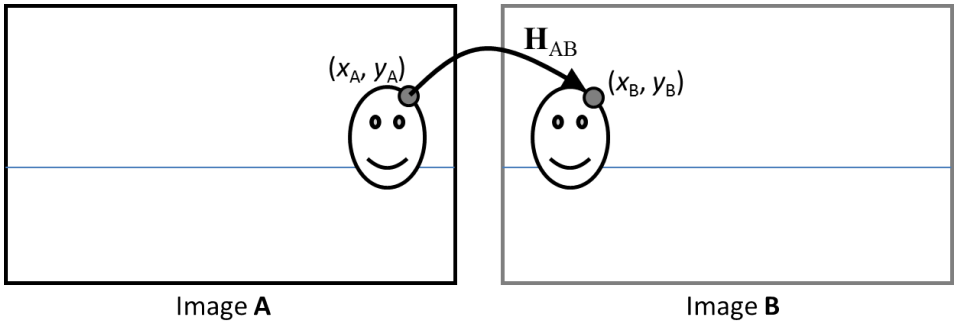
c. Homography

10. Recall that for each feature match we also have a distance between the two features, e.g. the SSD or ratio distance (in descriptor space, not (x,y) distance). Propose a way to improve on the RANSAC algorithm using these known distances.

11. Suppose we have a pinhole camera with a center of projection at the origin, and looking down the negative-*z* axis (i.e. in canonical camera coordinates). The distance from the camera to the projection plane is 20 units (recall the projection plane is parallel to the *x*-*y* plane). Which 2D point would a point in the scene at coordinates (32, 20, -40) project to in the image? Show your work. We've drawn a picture (in 2D) to get you started.

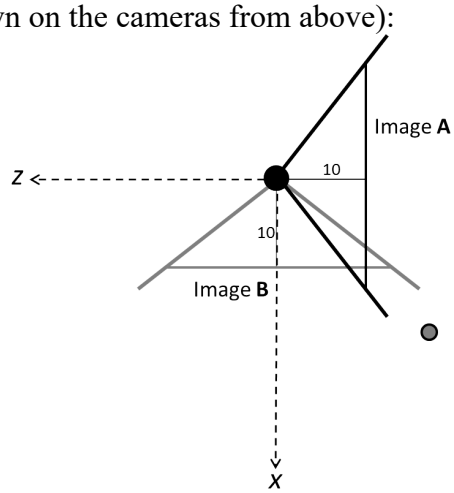


12. Suppose we capture two images from a camera undergoing a pure rotation (i.e., two images taken from the same center of projection) with a (wide-angle) perspective camera, so that the right side of Image **A** overlaps with the left side of Image **B**:



As described in class, these two images are related by a homography (we do not spherically project the images). This homography, $\mathbf{H}_{AB}$, is illustrated above. In this problem you will derive $\mathbf{H}_{AB}$ in terms of the camera parameters of images **A** and **B**.

In world coordinates, suppose that **A** has a center of projection at the origin, and is looking down the negative z-axis (i.e., it is in canonical form, with an identity rotation), and **B** has the same center of projection but is rotated 90 degrees about the y (up) axis, so that it is looking down the positive x-axis. Both cameras have a focal length of 10 units. This setup is illustrated below (from a viewpoint looking down on the cameras from above):



a) For each camera (**A** and **B**), what is its **K** (intrinsic) and **R** (rotation) matrix? Use an image coordinate system where the center of the image is the image origin. (Hint: each should be 3x3.)

$\mathbf{K}_A =$ $\mathbf{R}_A =$

$\mathbf{K}_B =$ $\mathbf{R}_B =$

b) Write the projection matrices for cameras **A** and **B**. Since the translation for both cameras is zero (as they are centered at the origin), you only need to write the first three columns of each projection matrix. Write each matrix first symbolically in terms of $\mathbf{K}_A$, $\mathbf{R}_A$, $\mathbf{K}_B$, and $\mathbf{R}_B$, then multiply out to get the full matrix.

$$\mathbf{\Pi}_A =$$

$$\mathbf{\Pi}_B =$$