Multiple choice (28 points)

1. **History (1 point).** The field of computer vision started:
   - A. ***As an undergraduate summer project in the 1960s***
   - B. As an undergraduate summer project in the 1990s
   - C. As a PhD student project in the 1960s
   - D. As a PhD student project in the 1990s

2. **Aperture (2 points).** Decreasing the aperture of a camera:
   - A. ***Increases the depth of field***
   - B. Decreases the depth of field
   - C. Does not affect the depth of field

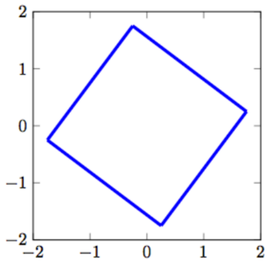
3. **Gaussian filter (2 points).** Consider a 1-D Gaussian with standard deviation of 2 pixels. Which of the following is the *most efficient* kernel that could capture at least 99% of this distribution?
   - A. Filter kernel of size 1×3
   - B. Filter kernel of size 1×7
   - C. ***Filter kernel of size*** $\boldsymbol{1\times13}$
   - D. Filter kernel of size 1×23

4. **SIFT pipeline (2 points).** Which of the following is *not* a part of the SIFT calculation pipeline?
   - A. Difference of Gaussians
   - B. Histogram of Gradients
   - C. ***Laplacian of Gaussians***
   - D. Orientation Histogram
   - E. Image Pyramid

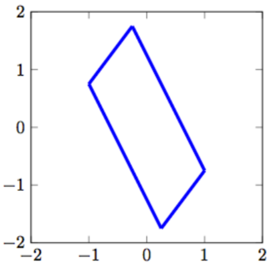
5. **SIFT robustness (2 points).** What is the SIFT feature least robust against?
   - A. Affine illumination
   - B. Rotation
   - C. Scale
   - D. ***Occlusion***

6. **RANSAC vs Least squares (3 points).** What are the benefits of RANSAC compared to the least squares method? *Circle all that apply.*
   - A. RANSAC is faster to compute in all cases
   - B. RANSAC has a closed form solution
   - C. ***RANSAC is more robust to outliers***
   - D. RANSAC handles measurements with small Gaussian noise better

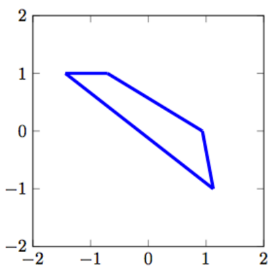
7. **Image Transformations (6 points).** The figures below show the outputs of applying one of the transformations on the right to a square with vertices at (1,1), (1, 0), (0, 0), and (0, 1). Circle the *most specific transformation* that could generate each output image.

(a) 

Projective Affine ***Similarity*** Rigid Body

(b) 

Projective ***Affine*** Similarity Rigid Body

(c) 

Projective Affine Similarity Rigid Body

8. **Affine transformation (3 points).** Which of the following always hold(s) under an affine transformation? *Circle all that apply.*

   A. ***Parallel lines will remain parallel*** .

   B. ***The ratio between two areas will remain the same*** .

   C. Perpendicular lines will remain perpendicular.

   D. The angle between two line segments will remain the same.

9. **Rotational invariance (3 points).** Which of the following representations of an image region are rotation invariant? *Circle all that apply.*

   A. ***Bag of words model with SIFT features***

   B. Spatial Pyramid Model with SIFT features

   C. A HOG template model

   D. Deformable Parts Model

10. **Bag of words (3 points).** Which of the following are drawbacks of bag of words (BOW) models? *Circle all that apply.*

   A. ***They don't capture spatial information***

   B. ***They are not ideal for solving detection problems***

   C. Creating BOW features is time consuming

   D. They are hard to understand because their inner workings are opaque

11. **Object categories (1 point).** As mentioned in lecture, most researchers working on object recognition would be familiar with the PASCAL VOC dataset. How many object categories are in PASCAL VOC?

    A. 1

    B. 10

    C. ***20***

    D. 101

Short answer (72 points)

Instructions: for each question, write your entire answer within the blank box immediately below the question. Unless otherwise specified, anything outside of the corresponding box will not be graded. You may use the margins or the last page of the exam for scratch space.

1. **Filters (8 points).** Give a 3x3 example of each of the following types of image filters:

   a. A brightening filter (only increases the image intensity)

   The middle entry is >1 ***and all other entries are*** 0***. Example:***

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

   b. A smoothing filter

   The entries are ≥ 0***, the entries add up to 1, and the filter is symmetric. Examples:***

$$\begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix} \begin{bmatrix} 0 & 0.1 & 0 \\ 0.1 & 0.6 & 0.1 \\ 0 & 0.1 & 0 \end{bmatrix}$$

   c. A sharpening filter

   The middle entry is >1***, all other entries are*** ≤ 0 ***(with some*** <0***), the entries add up to 1 and the filter is symmetric. Examples:***

$$\begin{bmatrix} 0 & -0.1 & 0 \\ -0.1 & 1.4 & -0.1 \\ 0 & -0.1 & 0 \end{bmatrix} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

   d. A filter for detecting a vertical edge

   Anything that approximates a dG/dx filter. The entries in the left (or right) column are >0***, and the entries on the other side have the corresponding negative values. The entries thus add up to 0. The filter is vertically symmetric. Examples:***

$$\begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

2. **Convolution and Filtering (7 points).**

Recall the definition of a convolution of an image f with filter g is:

$$(f * g)(u) = \sum_x f(u - x)g(x)$$

Prove that convolution with the unit impulse filter will leave the image unchanged.

Let f ***be the image and*** g ***be the unit impulse filter indexed by*** k, l***, where***

$$g[i, j] = \begin{cases} 1, & \textbf{\textit{if }} i = 0, j = 0 \\ 0, & \textbf{\textit{otherwise}} \end{cases}$$

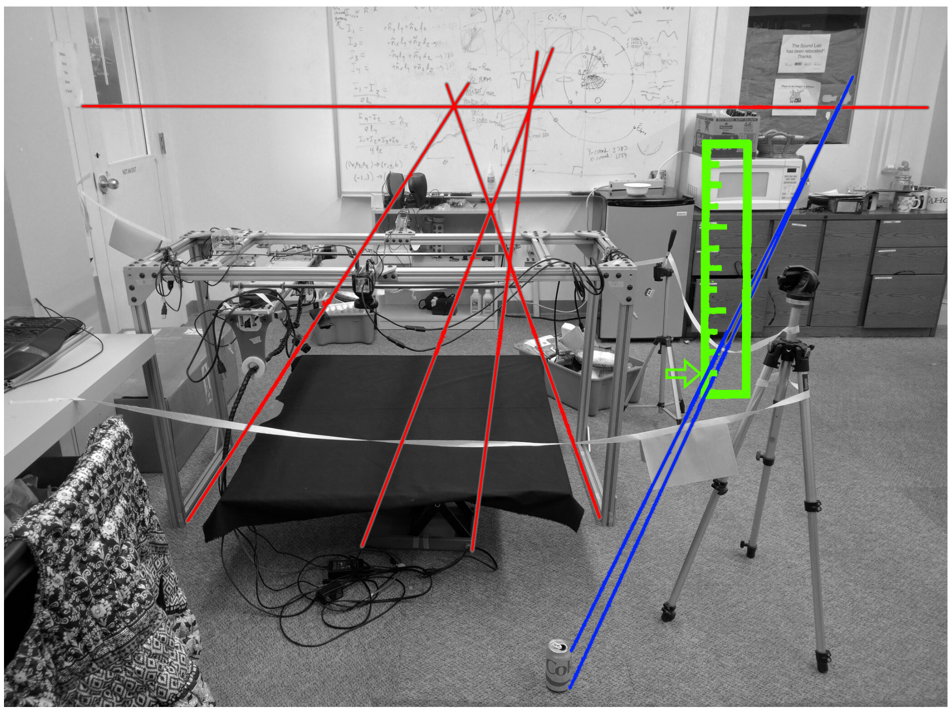
The output of convolving f ***with*** g ***is denoted*** $f * g$***, where***

$$\begin{aligned}
(f * g)[i, j] &= \sum_{k,l} f[i - k, j - l]g[k, l] \\
&= \sum_{k,l} f[i - k, j - l]g[k, l] \\
&= \sum_{k=0,l=0} f[i - k, j - l]g[k, l] + \sum_{k \neq 0 \textbf{\textit{ or }} l \neq 0} f[i - k, j - l]g[k, l] \\
&= f[i, j] + 0 \\
&= f[i, j].
\end{aligned}$$

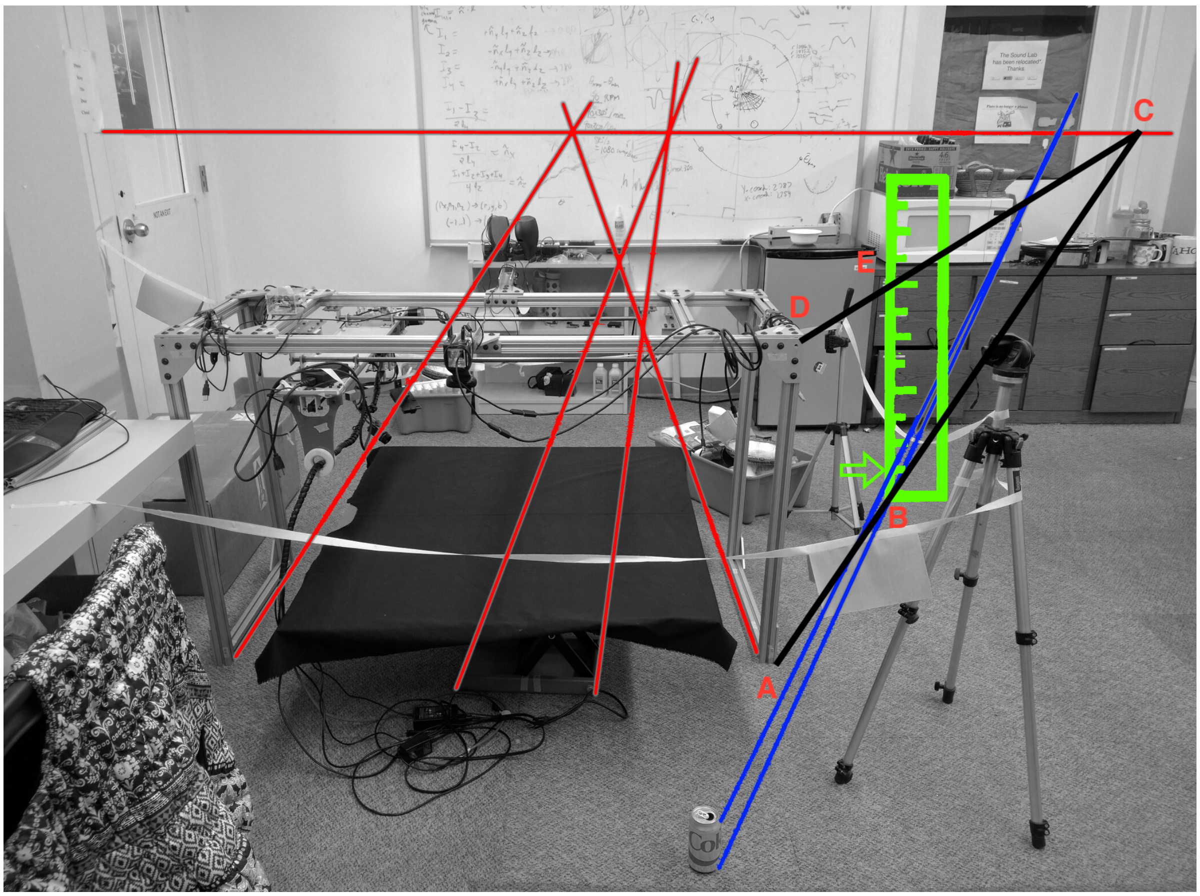
Therefore, convolution with the unit impulse filter will leave the image unchanged.

3. **Image Formation (8 points).** Your friend shows you the following image, hoping to find out exactly how tall the 3D scanner is (the scanner is the large metal frame). The scanner is custom-built, so you can't just look the answer up. Remembering your COS429 training, a quick internet search reveals that a coke can is 5 inches tall. How tall is the scanner? Show your work for full credit.

Instructions: for this question, there are two boxes (one on the image, one on the next page) which can be used for your answer.



We will measure the distance between point A and D as the scanner's height. The four non-horizontal red lines are two pairs of parallel lines on the ground (co-planar), so the line connecting both vanishing points is the horizon. We first shoot a ray from point A to point B, we name the intersection of the ray and the horizon as point C. Then we connect point C and D with a line, and the intersection of line CD and the edge of the green ruler (point E) shows the height of the scanner. From the question, we know that 5 inches corresponds to one interval on the ruler, and we can see there are 8 intervals on the segment BE. Since the height are proportional to the number of intervals, we know that the height of the scanner is 40 inches.



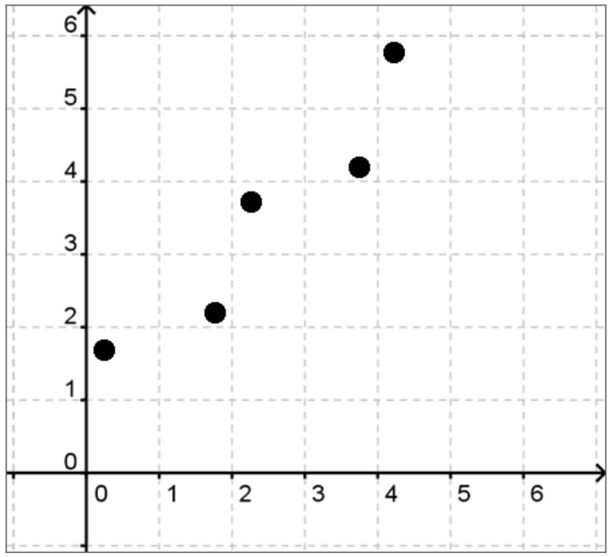
4. **Feature Detectors and Descriptors (6 points).** For each of the image locations A, B, and C, (they are marked with small blobs) what will the eigenvalues of the second moment matrix be? Describe in words how large each of the values will be and their relationship to one another.

A — ***Both eigenvalues will be similarly small.***

B — ***One eigenvalue will be small, and the other will be the large. The smaller eigenvalue is significantly smaller than the larger eigenvalue.***

C — ***Both eigenvalues will be similarly large.***

5. **Fitting, Hough Transforms, and RANSAC (20 points).** Suppose that we have a set of 2D point observations, and we want to fit a line to them. However, we already know the right slope: 1. Our observations are below.



a. What is the equation of our model? (2 points)

$$y = x + c$$

where c is the parameter

b. How many observations do we need to fit a model? (2 points)

1

c. Suppose (for the remaining parts of this question) we decide to apply the Hough transform to find the right model. How many models does each observation vote for? What shape does this take on in Hough transform space? (3 points)

Each observation votes for 1 model. The Hough transform space is 1-dimensional and each model is 1 point.

d. Suppose we use bins of size 2, with bins $[0, 2), [2, 4)$, and so on. What line will we fit? How many inliers and outliers will there be? (4 points)

$$y = x + 1$$

There will be 5 inliers and 0 outliers.

e. Suppose instead our bins are of size 1, with bins $[0, 1), [1, 2)$, and so on. What line will we fit? How many inliers and outliers will there be? (2 points)

$$y = x + 1.5$$

There will be 3 inliers and 2 outliers.

f. Discuss one way to make the Hough transform method more robust against rounding issues that arise with smaller bin sizes. (Hint: recall *splatting* from the image alignment lecture). (7 points)

An observation can vote for multiple bins, weighted according to its distance to the bin centers. The votes should be normalized though so each observation contributes the same total votes. For example, if the bins are [0,1), [1,2) then an observation of 0.9 can vote for bin [0,1) with weight 0.6 and for bin [1,2) with weight 0.4 (since it's 1.5x closer to the center of bin [0,1) than to the center of bin [1,2)).

6. **Recognition (23 points)**
Grading here is subjective but we looked for both correctness and diversity in the answers. Including specific examples to illustrate your responses is a great idea to help make sure the graders understand what you mean.

a. Briefly describe 3 reasons why template matching of RGB values does not work for object recognition in general. (6 points)

> 1. ***Recognizing an object with a single template is always difficult as this makes the matching highly dependent on the view angle and orientation.***
> 2. ***RGB values of the same object instance may vary a lot depending on factors such as illumination.***
> 3. ***Different instances of the object class may be different colors (for example, a yellow vs. a red car).***
> 4. ***RGB values of the target object class and other classes may be similar, as color is not a very discriminative feature (for example, red pens vs red pencils).***

b. Given an example of a car, why does matching SIFT keypoints to this example not work for recognizing other cars? Briefly describe 2 reasons. (4 points)

> 1. ***SIFT is good at finding local features in small regions. The local appearance would look different between two different car models, resulting in poor SIFT matches (for example, different shapes of headlights).***
> 2. ***The global structure and shape of two cars might be different (for example, the wheels might be at different distances from each other), making it difficult to build a model off of local SIFT matches.***

c. Give 1 example when a color histogram works well as an object descriptor. Briefly describe 3 situations when it does not work well. For each of the 3 situations, describe an approach discussed in class that was tried to improve on this weakness. (13 points)

A color histogram would work well when lighting and illumination are mostly constant, and the object of interest has a distinct color (for example, recognizing the single green plant in a yellow dessert).

Color histograms don't work well when:

1. ***Color alone is not a discriminative feature (for example, trying to identify the snowman object in a snow field)***
   Possible solution: use edge-based shape features such as HOG
2. ***The color pattern rather than just the color on the object is important (for example, finding the mug with black-and-white zebra-like stripes rather than the mug with black-and-white panda-like spots)***
   Possible solution: use a histogram of clustered SIFT features to encode the representation
3. ***The object's color changes depending on its viewpoint/orientation (for example, a sheet of white paper that has black writing on one side)***
   Possible solution: use a combination of histograms and accept any one as a valid match

END OF EXAM

PAGE INTENTIONALLY BLANK
Use for scratch work or for extra space