Midterm #2

November 27, 2018
Weight: 15% of final mark (150 points)
Time allotted: 90 minutes

ID number: ____________ **Name:** Dana

Instructions PRIOR to beginning the exam

a) Please **DO NOT** start the exam or flip the pages until instructed to do so.
b) Write your name and ID number in the spaces provided at the top of this page
c) Please place your **student identification card** on the desk for verification.
d) **No electronic aids** (phones / calculators / laptops / tablets / etc.) are allowed.
e) This is a closed book exam. **One single page of handwritten notes is allowed.** Notes collected at the end.

Instructions DURING the exam

a) Carefully read all instructions.
b) Observe the relative value of each question and **budget your time** accordingly.
c) Answer questions neatly in the space provided. You may do the questions in any order.
d) The back page of your exam is blank. Feel free to use it as scrap paper if you wish.
e) When finished, please hand in your entire exam paper to your instructor.

Grade Table (for teachers use only)

Question	Points	Score
1-10	50	
11	15	
12	20	
13	10	
14	10	
15	10	
16	10	
17	15	
18	10	
TOTAL	**150**	

Part 1 multiple choice questions: 50 points, 5 points each

NOTE: One, none or multiple choices could be true. Circle the answers that are correct.

1. In a regular camera with a lens:
    a. All objects are equally in focus
    b. Objects size is ambiguous
    c. Lines in the real-world project to lines in the image
    d. The image is inverted upside down

2. In a pinhole camera:
    a. All objects are equally in focus
    b. Objects size is ambiguous
    c. Lines in the real-world project to lines in the image
    d. The image is inverted upside down

3. In optic flow
    a. The motion on the image plane between two consecutive frames is calculated
    b. The motion with respect to a template (frame) is calculated
    c. Image constancy should hold
    d. Inter-frame motion has to be small
    e. For a small patch, spatial coherence in motion is assumed

4. In registration-based tracking
    a. The motion on the image plane between two consecutive frames is calculated
    b. The motion with respect to a template (frame) is calculated
    c. Image constancy should hold
    d. Inter-frame motion has to be small
    e. For a small patch, spatial coherence in motion is assumed

5. In the aperture problem
    a. Motion along the image gradient cannot be determined
    b. Motion along the image edge cannot be determined
    c. Motion perpendicular to the image gradient cannot be determined
    d. Motion perpendicular to the image edge cannot be determined

6. Computing the focal length of the camera can be done
    a. Using one image
    b. Using two images
    c. With the knowledge of the geometry of a 3D object
    d. Without the knowledge of the geometry of a 3D object
    e. Using a set of 2D-2D corresponding points
    f. Using a set of 2D-3D corresponding points

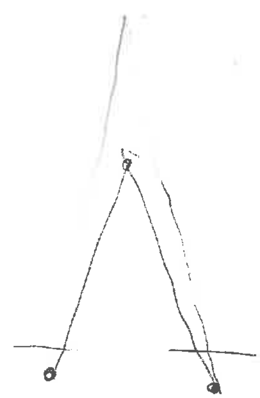
7. Computing the relative position (motion) of two images can be done
   a. Using one image
   b. Using two images
   c. With the knowledge of the geometry of a 3D object
   d. Without the knowledge of the geometry of a 3D object
   e. Using a set of 2D-2D corresponding points
   f. Using a set of 2D-3D corresponding points

2 views and cameras
→ F → E → R,t
1 view + geometry 3D
→ R,t from 3x4 P.

8. In projective geometry (under perspective projection)
   a. Lines map to lines
   b. Intersections are preserved
   c. Ratios of distances are preserved
   d. Parallel line intersect
   e. A rectangle maps to a parallelogram

9. In affine geometry (under affine transformations)
   a. Lines map to lines
   b. Intersections are preserved
   c. Ratios of distances are preserved
   d. Parallel line intersect
   e. A rectangle maps to a parallelogram

10. When reconstructing depth from a stereo view
   a. Precision of estimated depth increases when baseline increases
   b. Precision of estimated depth increases when baseline decreases
   c. With a fixed baseline, precision is inverse proportional to depth magnitude
   d. With a fixed baseline, precision is direct proportional to depth magnitude

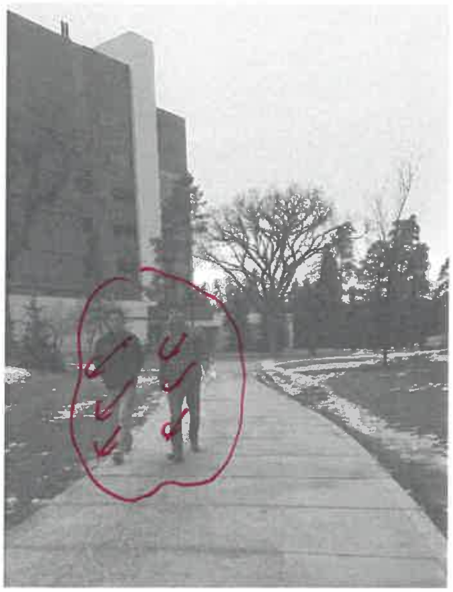


larger z - less precision.

Part 2 Dynamic vision: optic flow and tracking – **45 points**

Consider a camera mounted on a car facing the driving direction.

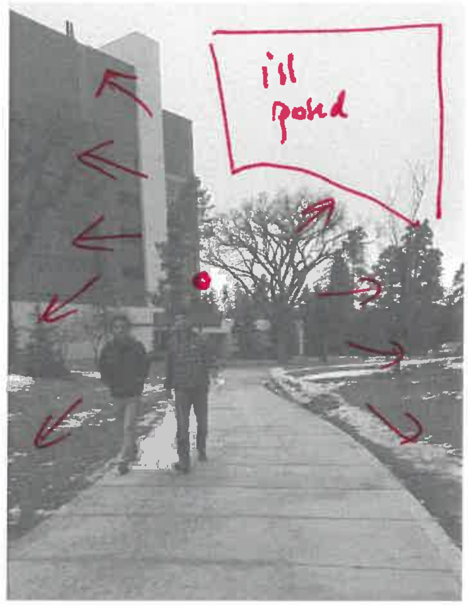
11. (15p) Sketch the flow field in the camera image in the following situations. Draw flow only for regions where it is well defined. Under each image, enumerate regions where the optic flow is ill-posed and explain what is the reason the optic flow cannot be calculated.



(a) Car is stopped

- motion flow only on object that move (people)
- no flow in all other regions

optic flow should be 0 but due to light change, wind etc. the image difference not exactly 0 → has to be made 0



(b) car moves forward

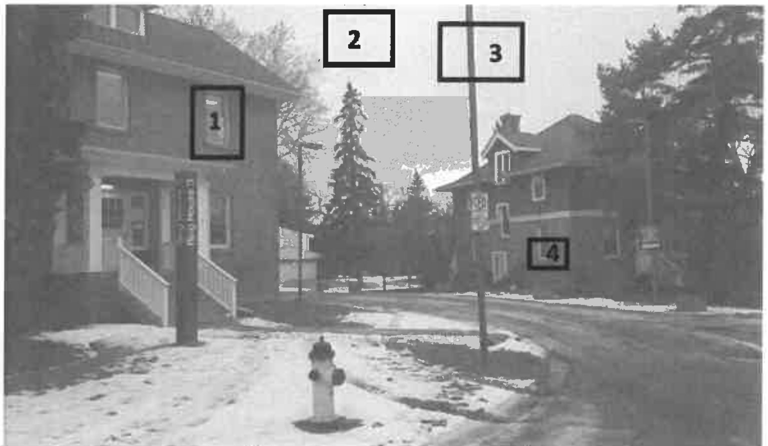
focus of expansion
motion flow vectors divergent from the focus of expansion
scale change → like zoom.

(c) What can you do to avoid ill-conditioned optic flow calculation in the areas where flow is not well defined? For which of the two cases (a) (b) is this most useful? *Hint : Think about your optic flow assignment.*

- to avoid ill-condition flow calculations in areas of no texture, no motion → image difference has to be made 0 if < thr.

(a) all areas where no motion (all except people) ← more useful here as most pix with 0.
(b) sky – no texture

12. (20p) Assume the car is moving forward. For each of the four patches depicted in the image, comment of the required tracking aspect



	Patch 1	Patch 2	Patch 3	Patch 4
Is tracking well posed? If not give one reason.	yes	no texture	aperture pb.	yes
What motion model would you use? How many DOF. Specify which DOF you are using (ex. translation, scale, rotation …)	at least 3DOF translation + scale could use affine or homography as patch large	X	translation only motion ⊥ pod can be detected	translation + zoom 3DOF
What would change if the patch is smaller?	if only inside window can loose features	X	same could become too small and lack features	smaller might only do 2DOF
What would change if the patch is bigger?	same better tracking if more features	if the tree is included tracking can become stable	same unless very big and house is included.	if larger we might need to increase DOF

13. (10p) Tracking

(a) Enumerate at least three main causes of errors in tracking. Write a possible solution to solve each of them.

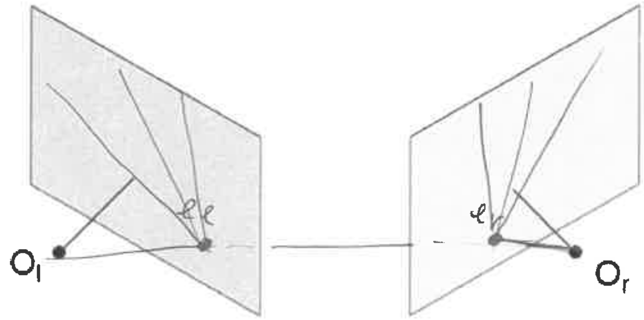
- large motion – violates small motion assumption → image pyramid can fix this
- light change / template change – violates image constancy → adaptive template, make robust to light changes
- no texture → larger patches

(b) If you track a patch of 100x100 pixels, tracking could be slow. Do we need all pixels? Can you suggest some way(s) to reduce the computational complexity?

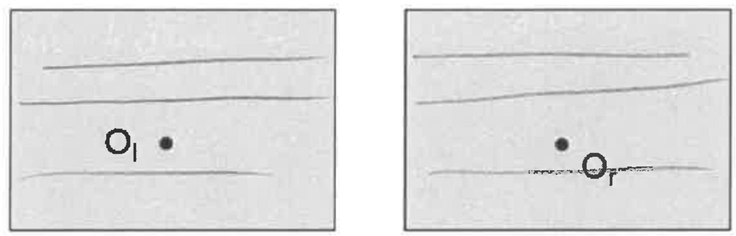
– we could downsample (take every 2-nd pixel like in the assignment implementation)

Part 3 Geometry– 55 points

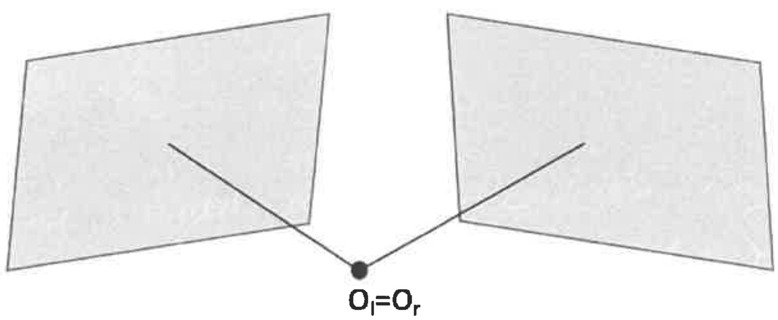
14. (10p) Draw the **epipoles** and **epipolar lines** for the four cases below.



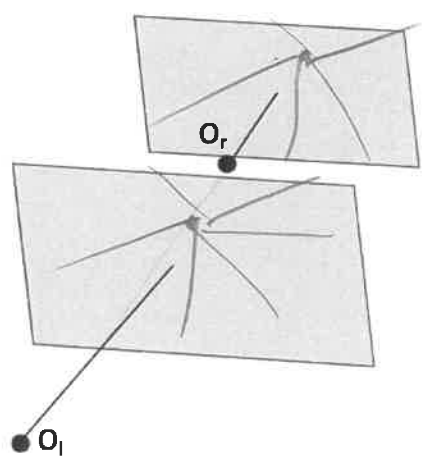
epipoles in the image
epipolar lines converge to epipoles.



stereo
-epipolar lines // and aligned with rows.
-epipoles at ∞

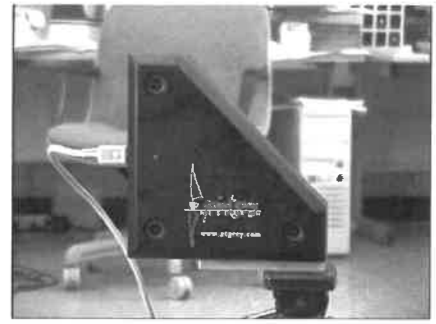


camera rotated around optical center
=> no epipoles, no epip. lines
cannot calculate depth



epipole in same location.
epipolar lines radiate from epipole

15. (10p) Consider the trinocular device from the image on the right.

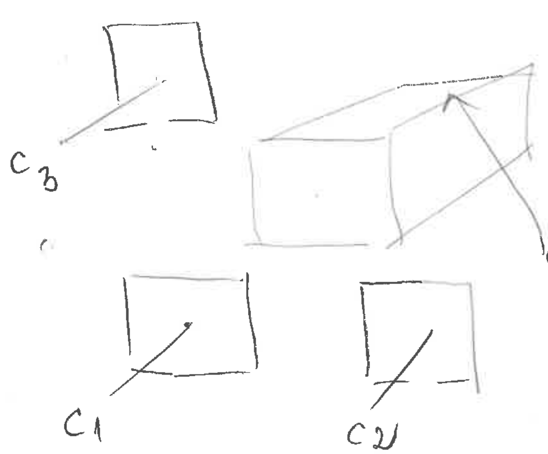


What do the 3 cameras bring in addition to a traditional stereo rig on the left?

3 cameras increase accuracy
fov increases – if some points have no overlap in 2 cameras might have overlap (common features) in the other 2 pairs.

Think about an object that is in a position such that depth cannot be recovered by the 2 horizontal cameras but can be recovered with the two vertical cameras. Draw a figure that illustrates the viewing situation.

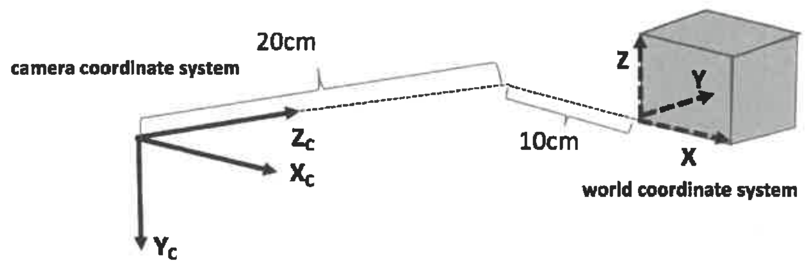
If an object is aligned with ~~z~~ axis ~~#~~ C_1 & C_2
→ large error in depth that can be corrected by C_1 that has a better angle



depth has errors here if estimated in C_1, C_2

Better estimated from $C_1 C_3$ or $C_2 C_3$

A camera is placed on a table. We define a standard camera coordinate system that has z-axis along the optical axis and x,y axis aligned with the image plane. A cube is placed at about 20cm in front of the camera, at the same height and about 10cm to the right of the optical axis. The cube has the front face parallel with the image plane. The world coordinate system is defined such that the z axis points vertically upwards and the origin is in the left-bottom corner of the cube like in the figure below.



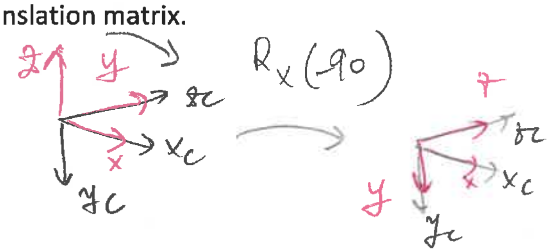
16. (10p) Write the rotation R and translation T that will align the world coordinate system with the camera coordinate system such that the projection equation holds

$$p = KRTP$$

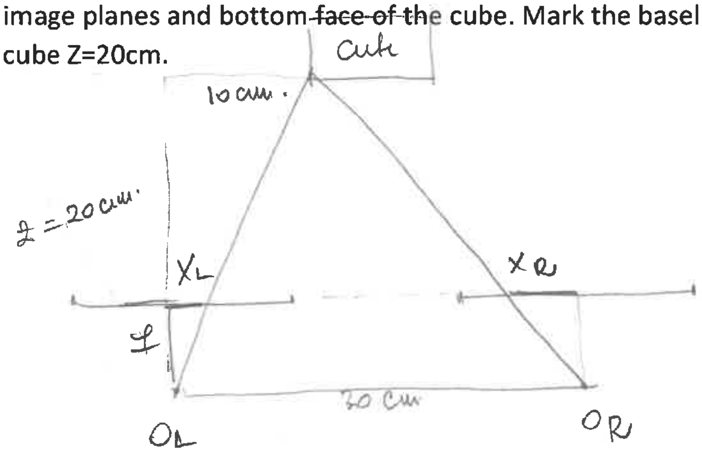
where p are the homogeneous coordinates of the image point (on the image plane), P are the homogeneous coordinates of the 3D point defined in the world coordinate system and K is the 3x3 camera matrix. R is a 4x4 rotation matrix and T is a 4x4 translation matrix.

1) translation. $T\begin{pmatrix}10\\20\\0\end{pmatrix}$

2) $R_x(90)$



17. (15p) If a second camera is placed on the right of the first camera at 30cm such that the two image planes are parallel with the line that connects the two optic centers (standard stereo configuration). Note that the two image planes are also parallel with the front face of the cube.
   (a) Make a 2D drawing (X_CZ_C plane) of this configuration, showing the two camera centers, the two image planes and bottom ~~face of the~~ cube. Mark the baseline B=30cm and the depth to the cube Z=20cm.



(b) Derive the focal length of the camera **f** if you know the image coordinates of the bottom-left corner of the cube in the two images (x_L, x_R). Use the stereo equation (with known baseline B and depth Z). Use your previous diagram to help. What are the units of **f**?

$$\frac{f}{Z} = \frac{X_L - X_R}{B} \Rightarrow f = \frac{Z(X_L - X_R)}{B}$$

units = pixels

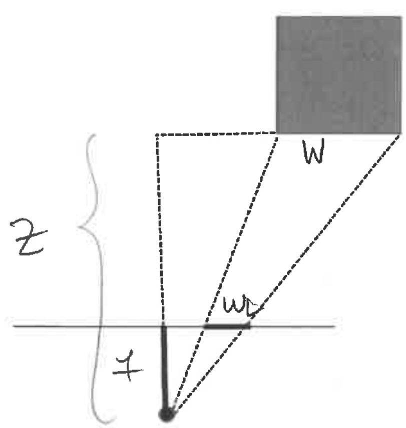
(c) What is the form of the 3x3 camera matrix K (from point 15) knowing f=800 and the image has dimension of 640x480 pixels (width,height).

$$K = \begin{pmatrix} 800 & 0 & 320 \\ 0 & 800 & 240 \\ 0 & 0 & 1 \end{pmatrix}$$

$$K = \begin{pmatrix} f & 0 & u_c \\ 0 & f & v_c \\ 0 & 0 & 1 \end{pmatrix}$$

u_c, v_c = center of the image

18. (10p) Calculate the length of the edge of the cube **W** knowing the length of the projection of the bottom edge on the left camera plane $\mathbf{w_L}$. Use the diagram below (XZ projection). On the diagram mark the following entities: focal length **f**, depth to the cube **Z**, the 3D width of the cube **W** (cube edge length) and the length of the image projection of the bottom edge **w**.



first(left) camera

$$\frac{f}{Z} = \frac{w_L}{W} \Rightarrow W = \frac{Z \cdot w_L}{f}$$