FACULTÉ
DES SCIENCES
ET TECHNIQUES
SAINT-ÉTIENNE

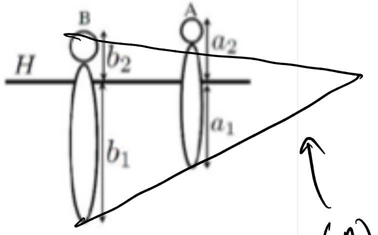
Master MLDM and 3DMT - Computer Vision course

Exam March 2019 - 2h without documents

(6 parts with a total of 15 questions accounting for 25 points (+ 4 bonus points), the exam will be scored for 20 points)

Part 1 *(5 points + 4 bonus points – 30 min)*: **Comparing heights**

Consider the image of person A and person B standing on the ground plane (see figure below), as taken by a perspective camera of focal length f. H is the horizon line (y=0). a_1 is the distance between the A's toes and the horizon, a_2 is the distance between A's head and the horizon in units of pixels. Similarly for person B. h is the tall of person A. Distances a_1, a_2, b_1, b_2, are expressed in the camera plane. Suppose A's height is h feet.



Question 1 *(3 points)*: From this Figure, who is taller? Briefly explain why? (A)

Question 2 *(2 points)*: How many feet above the ground is the camera? Give the equation which relies the height h_c and the following parameters: a_1, a_2 and h. Briefly explain why?

$h_c = \frac{a_1 h}{a_1 + a_2}$ $\qquad$ $h_c = \frac{b_1}{b_1 + b_2} \times h_b$

Hint: see Annex

Question 3 *(bonus question, 2 points)*: How tall is person B (in feet)? Give the equation which relies the height h_b and the following parameters: a_1, a_2, b_1, b_2, and h. Briefly explain why?

$h_b = \frac{b_1 + b_2}{b_1} \cdot h_c = \frac{b_1 + b_2}{b_1} \cdot \frac{a_1 h}{a_1 + a_2} = \frac{h a_1 (b_1 + b_2)}{b_1 (a_1 + a_2)}$

Question 4 *(bonus question, 2 points)***:** What is the distance (along the z-axis) from the camera to person B (in feet)? Give the equation which relies the height of person B and the following parameters: a_1, a_2, b_1, f, and h. Briefly explain why?

$$\frac{x}{z} = -f\frac{X}{Z}$$

$$\Rightarrow z = \frac{Zx}{fX} \Rightarrow \frac{b_1+b_2}{f} = \frac{h_b}{Z}$$

$$\Rightarrow Z = \frac{f h_b}{b_1+b_2}$$

Part 2 *(4 points – 20 min)***: Short answers**

Question 5 *(1 point)***:** Express the point where 2-D parallel lines intersect in homogeneous coordinates (assume any notation).

$Ax+By+C=0$

$Ax+By+D=0$

$x \to \frac{x}{w}$

$y \to \frac{y}{w}$

$Ax+By+Cw=0$

$Ax+By+Dw=0$

$(C-D)w=0$

$\to w=0$

Question 6 *(1 point)***:** How many degrees of freedom are there in perspective projection transformation between two images?

8

Question 7 *(1 point)***:** Considering that:

$$\mathbf{E}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \mathbf{M}\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

What are the coefficient of the matrix M in case of orthographic projection?

Hint: see Annex

Question 8 *(1 point)***:** A projective transformation between two images X and X' can be modelled by transforming each point p in one image to a point p' in a new image using p'= H p, where H is the homography. Briefly explain, why we can't do the same thing with the Fundamental matrix F to transform each point p in the left image to a point p' in the right image from the transform p' = Fp?

Hint: see Annex

(If we use fundamental matrix, the pt. in one image maps in epipolar line in other pt.)

Part 3 *(2 points – 10 min)***: Estimating depth**

Question 9 *(2 points)***:** Assume that humans can discriminate disparity difference of 6". Further assume that the distance between the eye (b) is 10 cm. What is the depth resolution (d) that humans can discriminate, when an object is fixated at a depth (Z) of:

1. 10 cm (in this case $Z \approx b$; $\tan(F_1) = b/2Z$ and $\tan(F_2) = b/2(Z+d)$)
2. 1 m (in cases 2 to 4, we can assume that $Z >> b$; $\tan(F_1 - F_2) = bd/2Z^2$)

3. 10 m
4. 100 m

$$\tan(f_1 - f_2) = \frac{\frac{b}{2Z} - \frac{b}{2(Z+d)}}{1 + \frac{b^2}{4Z(Z+d)}}$$

Hint: 1° = 60 min arc = 3600 sec arc (i.e. 3600”); 1 rad = 57° 17’ 44”

$$2.91\times10^{-5} = \frac{10\times d}{8\cdot10\,(10+d)+2\cdot100}$$

$$\Rightarrow 2.9\times10^{-5} = \frac{d}{8(10+d)+20}$$

$$\Rightarrow 34377 = \frac{80}{d} + 8 + \frac{20}{d}$$

$$\frac{\frac{b(Z+d) - bZ}{2Z(Z+d)}}{\frac{4Z(Z+d)+b^2}{4Z(Z+d)}} = \frac{2bd}{4Z(Z+d)+b^2}$$

$$\frac{b\times 10}{2\times100\times100}$$

$$\Rightarrow b = 20\times\tan(f_1-f_2) = 0.058\text{ cm}$$

Part 4 *(5 points – 20 min)*: **Perspective**

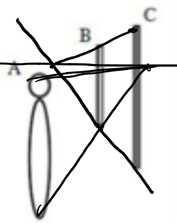
Question 10 *(1 point)*: Which scenario is most suitable for scale orthographic camera model?

1. Describe the movement of a car in front of another car.
2. Image formation using a laptop.
3. Describe the movement of a plane at very high altitude.

Question 11 *(2 points)*: Consider the image of person A standing on the ground plane and two vertical utility poles B and C on the same ground plane (see image below) as taken by a perspective camera of focal length f. Consider that pole B and C have the same height. Determine the horizon line and draw it on the image below.

$$\Rightarrow 34369.4f = \frac{100}{d}$$

$\Rightarrow d =$



Question 12 *(2 points)*: From the image, is the person A taller than the pole B? Why?

Hint: see Annex

Smaller

Part 5 *(3 points – 10 min)*: **Stereo correspondence**

Question 13 *(3 points)*: A stereo system is used on a flying vehicle. It consist of two ccd cameras, each having 512x512 detectors on a 1 cm^2. The lenses used have 16 mm focal lengths (f) (and the focus is fixed at the infinity). For corresponding points (u_1, v_1) in the left image and (u_2, v_2) in the right image, v_1=v_2 as the x-axes in the two image planes are parallel to the epipolar lines. The vergence is adjusted so that disparity u_2-u_1 of points at infinity is zero in the two images derived from the cameras. The baseline (b) between the cameras is 1 foot. If the nearest range to be measured is 16 feet (Z), what is the largest disparity that will occur (in pixels)? What is the range resolution, due to the pixel spacing, at 16 feet? What range corresponds to the disparity of one pixel?

Disparity

$D = \frac{fb}{Z_c}$ ← focal length, ← baseline, ← range

disparity

left image

$x_l = \frac{fx}{Z}$

$x_r = f\left(\frac{x-b}{Z}\right)$

$x_l - x_r = \frac{fx}{Z} - \frac{f(x-b)}{Z} = \frac{fb}{Z}$

Hint: 1 feet = 30.48 cm.
Hint : definition of vergence in annex

Part 6 *(6 points – 20 min)*: **Calibration**

We seek to calibrate a rigid body camera with a planar calibration object. The object possesses M distinct features. The location of those features in the object coordinate frame are unknown (and they are not arranged in a checkerboard). Let be K the number of images acquired and M the number of features on the planar calibration object.

In answering the questions below, you might want to use the following notation/equations

---

$$\begin{pmatrix} \tilde{X}^C \\ \tilde{Y}^C \\ \tilde{Z}^C \end{pmatrix} = R \begin{pmatrix} X^W \\ Y^W \\ Z^W \end{pmatrix} + \begin{pmatrix} T_X \\ T_Y \\ T_Z \end{pmatrix}$$

$$R = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \cos\varphi & 0 & \sin\varphi \\ 0 & 1 & 0 \\ -\sin\varphi & 0 & \cos\varphi \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{pmatrix}$$

$$\begin{pmatrix} x_{\text{im}} \\ y_{\text{im}} \end{pmatrix} = \left\lfloor \begin{pmatrix} \frac{f}{s_x} \frac{\tilde{X}^C}{\tilde{Z}^C} + o_x + 0.5 \\ \frac{f}{s_y} \frac{\tilde{Y}^C}{\tilde{Z}^C} + o_y + 0.5 \end{pmatrix} \right\rfloor$$

Question 14 *(2 points)*: What are the free (i.e. intrinsic and extrinsic) parameters that can be recovered through the calibration process? Compute the number of unknown parameters defined above.

Question 15 *(4 points)*: Let us assume that there are 2M parameters for the unknown feature locations $\{X^W_j; Y^W_j\}$ on the planar calibration object and that $Z^W_j = 0$ for all features j. So, we have to add the object parameters $\{X^W_j; Y^W_j\}$ to the set of free parameters. How many parameters have to be estimated to calibrate the camera from this calibration object? How many constraints are being provided by each image of the calibration pattern? What are the lower bounds for M and K? Provide exact formulae, one of the form K 3… and one of the form M 3… For K = 4 images, what is the minimum M? For M = 4 features, what is the minimum K?

Annex

$$\begin{pmatrix} \frac{f}{S_x Z_c} & 0 & 0 & O_x + 0.5 \\ 0 & \frac{f}{S_y Z_c} & 0 & O_x + 0.5 \end{pmatrix} \begin{pmatrix} \tilde{X}_c \\ \tilde{Y}_c \\ \tilde{Z}_c \\ 1 \end{pmatrix}$$

$X_1^w = 0$
$Y_1^w = 0$
$Z_1^w = 0$

$\therefore (2M+6K+4-3)$

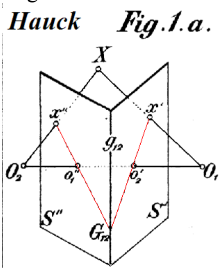
From Wikipedia: $(2M+6K+1)$

$\therefore 2KM \geq (2M+6K+1)$

$\Rightarrow 8M \geq 2M+24+1$

$\Rightarrow 6M \geq 25 \Rightarrow M \geq 5$

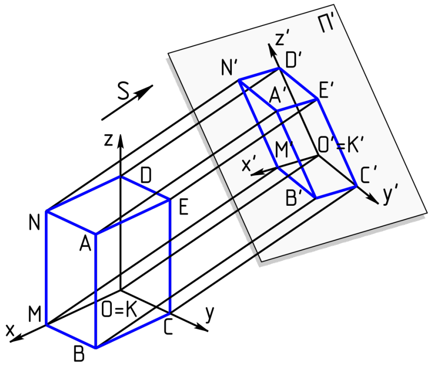
In the field of computer vision, any two images of the same planar surface in space are related by a **homography** (assuming a pinhole camera model). This has many practical applications, such as image rectification, image registration, or computation of camera motion—rotation and translation—between two images. Once camera rotation and translation have been extracted from an estimated homography matrix, this information may be used for navigation, or to insert models of 3D objects into an image or video, so that they are rendered with the correct perspective and appear to have been part of the original scene.



Geometrical setup for homography: stereo cameras O_1 and O_2 both pointed at *X* in epipolar geometry.

From Wikipedia:

Orthographic projection (sometimes **orthogonal projection**) is a means of representing three-dimensional objects in two dimensions. It is a form of parallel projection, in which all the projection lines are orthogonal to the projection plane, resulting in every plane of the scene appearing in affine transformation on the viewing surface. The obverse of an orthographic projection is an oblique projection, which is a parallel projection in which the projection lines are *not* orthogonal to the projection plane.



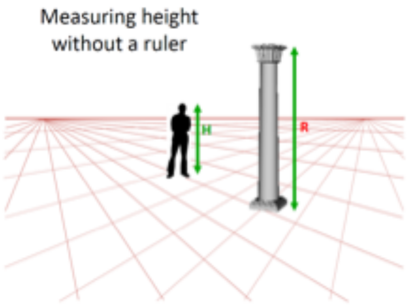
A simple orthographic projection P onto the plane $z = 0$ can be defined by the following matrix:

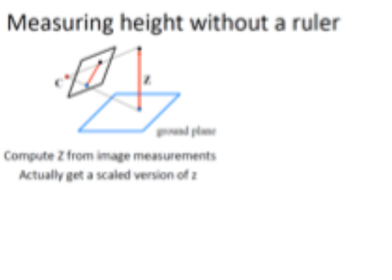
$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

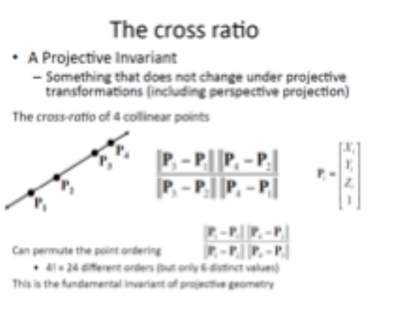
For each point $v = (v_x, v_y, v_z)$, the transformed point would be:

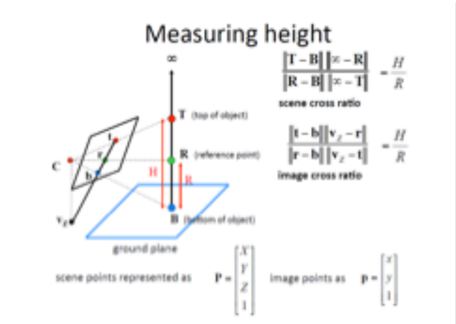
$$Pv = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} = \begin{bmatrix} v_x \\ v_y \\ 0 \end{bmatrix}$$

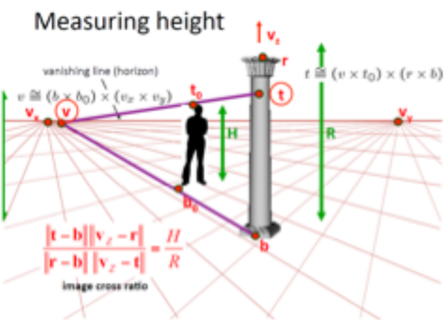
From http://www.cs.cornell.edu/courses/cs4670/2015sp/lectures/lec19_svm2_web.pdf

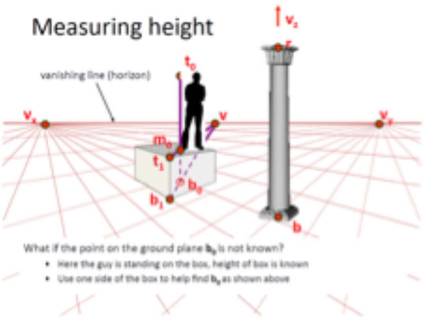








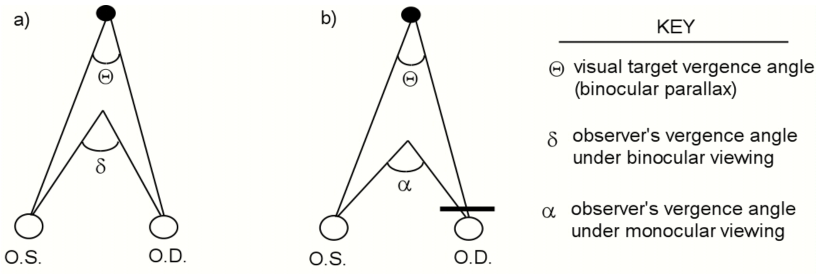




From http://schorlab.berkeley.edu/passpro/oculomotor/html/lab7.html

Under binocular viewing, subjects generally do not gaze directly at the visual target. Fixation disparity is defined as the difference between the target **vergence** angle (binocular parallax) and the ocular **convergence** angle during binocular fixation, as shown in Fig. 1. Fixation disparity occurs in the presence of binocular feedback, so it is a closed-loop error. It is measured under the assumption that the fixation target remains fused.

Phoria is defined as the difference between binocular parallax and the ocular convergence angle during monocular fixation (i.e., when one eye is occluded), as shown in Fig. 1. Occluding one eye dissociates binocular vergence by eliminating feedback from binocular retinal image disparity, so it is an open-loop vergence error. The phoria indicates the position of rest of the eyes. The magnitude of the phoria is proportional to the magnitude of the fixation disparity.


a)
Θ
δ
O.S.
O.D.
b)
Θ
α
O.S.
O.D.
KEY
Θ visual target vergence angle (binocular parallax)
δ observer's vergence angle under binocular viewing
α observer's vergence angle under monocular viewing