

# Midterm Exam Solutions

CS 682: Computer Vision (J. Košecká)

April 13, 2004

**HONOR SYSTEM:** This examination is strictly individual. You are not allowed to talk, discuss, exchange solutions, etc., with other fellow students. Furthermore, you are only allowed to use the book and your class notes. You may only ask questions to the class instructor. Any violation of the honor system, or any of the ethic regulations, will be immediately reported according to George Mason University honor court.

1. (15) **Filtering.** The image below is an image of a 3 pixel thick vertical line.

- (a) Show the resulting image obtained after convolution of the original with the following approximation of the derivative filter  $[-1, 0, 1]$  in the horizontal direction. How many local maxima of the filter response do you obtain ?

```
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
```

**Solution** The response of each row after convolution with the above filter will be

```
0 0 1 1 0 -1 -1 0 0
```

There are two local extrema at 1 and -1.

- (b) Suggest a filter which when convolved with the same image would yield a single maximum in the middle of the line. Demonstrate the result of the convolution on the original image.

```
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 0 0
```

**Solution** Convolving the image with the following filter will yield single maximum in the middle of the line  $g = [1, 2, 1]^T$ .

- (c) What is a difference between Gaussian smoothing and median filtering? How would you decide to use one vs. another ?

Median filter is more used for noise removal, more suitable for salt and pepper noise. Gaussian is used for smoothing and its better when the noise is not spiky.

2. (10) **Image Motion.** Consider an image motion model (related to affine and translational models considered in the class) which properly models the following local changes of image appearance: image translation  $d = [d_1, d_2]^T$ , change in image contrast  $\lambda$ , intensity offset  $\delta$  and simple isotropic expansion modelled by a parameter  $a$ . Derive a method for estimating the unknown parameters of this model and discuss conditions when such method would fail.

**Solution** The image motion model which we will consider, is derived from modified version of the brightness constancy constraint introduced for the affine motion model and has the following form:

$$I(\mathbf{x}, t) = \lambda I(A\mathbf{x} + d, t + dt) + \delta$$

Where matrix  $A$  is:

$$A = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$$

with  $a$  being the expansion parameter and  $d = [d_1, d_2]^T$ . Similarly as in the pure translational case we will try to expand the RHS of the above constraint using Taylor series expansion. The RHS can be written as function of additional parameters of  $\lambda$  and  $\delta$  denote by

$$\begin{aligned} g(x + dx, y + dy, t + dt, \lambda + d\lambda, \delta + d\delta) &= \lambda I(x + ax - x + d_1, y + ay - y + d_2, t + dt) + \delta \\ &= \lambda I(x + dx, y + dy, t + dt) + \delta \end{aligned} \quad (1)$$

with  $dx = ax - x + d_1$  and  $dy = ay - y + d_2$ . Now expanding the LHS above at  $g(x, y, t, 1, 0)$  we get

$$g(x + dx, y + dy, t + dt, \lambda + d\lambda, \delta + d\delta) = g(x, y, t, 1, 0) + \frac{\partial g}{\partial x} dx + \frac{\partial g}{\partial y} dy + \frac{\partial g}{\partial t} dt + \frac{\partial g}{\partial \lambda} d\lambda + \frac{\partial g}{\partial \delta} d\delta + \epsilon.$$

Now evaluating the partial derivatives at the point of expansion and keeping only the first order terms we get

$$g(x, y, dt, \lambda, \delta) = f(x, y, t) + \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt + \frac{\partial I}{\partial \lambda} (\lambda - 1) + \frac{\partial I}{\partial \delta} (\delta - 0).$$

Substituting the RHS back to the original equation and substituting back for  $dx$  and  $dy$  we will get the following version of the brightness constancy constraint

$$\frac{\partial I}{\partial x} ((a - 1)x + d_1) + \frac{\partial I}{\partial y} ((a - 1)y + d_2) + \frac{\partial I}{\partial t} dt + I\lambda_e + 1\delta = 0$$

where  $\lambda_e = \lambda - 1$ . Separating the unknowns and the measurements and denoting  $a' = a - 1$ , we can rewrite the above constraint in the following form

$$[xI_x + yI_y, I_x, I_y, I, 1]^T [a', d_1, d_2, \lambda, \delta] = -I_t$$

where  $I_x = \frac{\partial I}{\partial x}$ ,  $I_y = \frac{\partial I}{\partial y}$ ,  $I_t = \frac{\partial I}{\partial t}$ . The least squares solution of the unknowns given above model follows in the same way as purely affine (or translational case). More details on page 383 of the book or in Chapter 4).

3. (15) **Paracatadioptric Cameras.** A paracatadioptric camera combines a paraboloidal mirror of focal length  $1/2$  and focus at the origin, with an orthographic projection. The equation of such paraboloidal mirror is given by

$$Z = \frac{1}{2}(X^2 + Y^2 - 1). \quad (2)$$

Therefore, the projection (image)  $\mathbf{x} = (x, y, 0)^T$  of a 3-D point  $q = (X, Y, Z)^T$  is obtained by intersecting a parameterized ray with the equation of the paraboloid to yield  $\mathbf{b}$  (see Figure 1), and then orthographically projecting  $\mathbf{b}$  onto the image plane  $Z = 0$ .

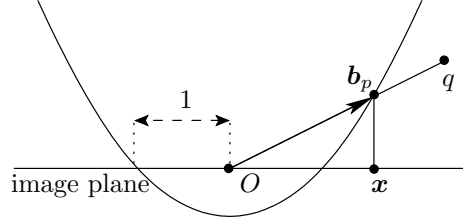


Figure 1: Showing the projection model for paracatadioptric cameras, and the back-projection ray  $\mathbf{b}$  associated with image point  $\mathbf{x}$ .

- (a) Show that the image of  $q = (X, Y, Z)^T$  is given by

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{-Z + \sqrt{X^2 + Y^2 + Z^2}} \begin{bmatrix} X \\ Y \end{bmatrix} = \frac{1}{\lambda} \begin{bmatrix} X \\ Y \end{bmatrix}. \quad (3)$$

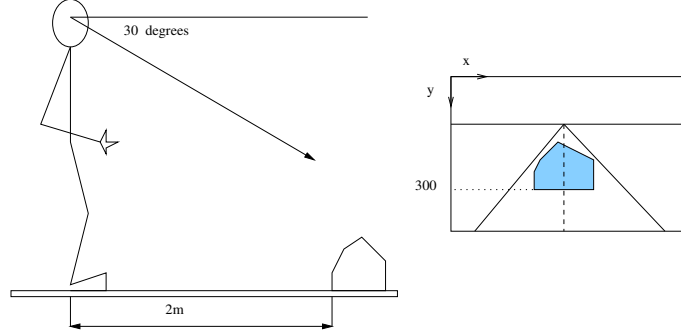
- (b) The back-projection ray  $\mathbf{b} \in \mathbb{R}^3$  is a ray from the optical center in the direction of the 3D point  $q \in \mathbb{R}^3$  being imaged (see Figure 1). Show that  $\lambda \mathbf{b} = q$ , where  $\mathbf{b} = (x, y, z)^T$  and  $z = (x^2 + y^2 - 1)/2$ .
- (c) Given two views  $\{(\mathbf{x}_1^i, \mathbf{x}_2^i)\}_{i=1}^N$  related by a discrete motion  $(R, T) \in SE(3)$ , can one apply the 8-point algorithm to the corresponding back-projection rays  $\{(\mathbf{b}_1^i, \mathbf{b}_2^i)\}_{i=1}^N$  to compute  $R$  and  $T$ ?

### Solution

- (a)-(b) The intersection of the ray from the optical center in the direction of the 3D point  $q$  with the paraboloid is a point  $\mathbf{b} = (x, y, z)^T = q/\lambda$ , where  $\lambda$  is a scale factor. Since  $\mathbf{b}$  is in the paraboloid, we have  $z = (x^2 + y^2 - 1)/2$ , hence  $Z/\lambda = ((X^2 + Y^2)/\lambda^2 - 1)/2$ . After multiplying by  $2\lambda^2$  and rearranging the terms we get  $\lambda^2 + 2Z\lambda - (X^2 + Y^2) = 0$ , from which  $\lambda = -Z \pm \sqrt{X^2 + Y^2 + Z^2}$ . Since  $\lambda > 0$ , then we have  $\lambda = -Z + \sqrt{X^2 + Y^2 + Z^2}$ .
- (c) We have  $q_2 = Rq_1 + T$ , hence  $\lambda_2 \mathbf{b}_2 = \lambda_1 R\mathbf{b}_1 + T$  and  $\mathbf{b}_2^T \widehat{T} R \mathbf{b}_1 = 0$ . Therefore, one can apply the 8-point algorithm directly to the back-projection rays  $\{(\mathbf{b}_1^i, \mathbf{b}_2^i)\}_{i=1}^N$  to compute  $R$  and  $T$ . The only difference is that the third entry of each  $\mathbf{b}$  need not be equal to 1, since the back-projection rays are obtained from the given images  $\{(\mathbf{x}_1^i, \mathbf{x}_2^i)\}_{i=1}^N$  by setting the third entry to  $z = (x^2 + y^2 - 1)/2$ . Notice that one may as well divide each  $\mathbf{b}$  by  $\mathbf{b}_z$ , and then consider  $\mathbf{x} = \mathbf{b}/\mathbf{b}_z$  as perspective images. Then one can apply the 8-point algorithm as usual.

4. (10) **Perspective projection.** Consider a person standing on the road viewing the road at the viewing angle  $\alpha$ .

- (a) How would you compute the viewing angle, providing that you can observe the  $y$  image coordinate of the horizon<sup>1</sup> line?
- (b) Suppose that computed viewing angle is  $30^\circ$  and that there is an obstacle in front the person at the distance of 2 meters from the feet. Consider that the parameters of the imaging system can be well approximated by a pinhole camera where the resulting image is of resolution  $400 \times 400$ , the focal length is  $f = 30$  and the image of the projection center is the center of the image. The  $y$ -coordinate of the obstacle in the image is 300 pixels. How tall is the person ?



**Solution** a) The coordinate of the horizon is related to the coordinate of the vanishing point, which is an intersection of the two parallel lines in the ground plane. I assume that  $y$ -axis is pointing downwards,  $z$ -axis if to the right and  $x$ -axis out of plane. The coordinates of the vanishing point are

$$x = \frac{X_c + \lambda v_1}{Z_c + \lambda v_3} \quad \text{and} \quad y = \frac{Y_c + \lambda v_2}{Z_c + \lambda v_3}$$

where  $v_c = [v_1, v_2, v_3]^T$  is the direction vector of a line in the camera coordinate frame and  $\mathbf{X} = [X_c, Y_c, Z_c]^T$  is the base point of the line. Consider two lines in the world coordinate frame which lie in the ground plane with direction vector  $v_w = [0, 0, 1]^T$ . Then the same direction vector in the camera frame will be  $v_c = [0, -\sin \alpha, \cos \alpha]$  (see the relationship between the camera and world coordinate frame in part b). Suppose that  $y'$  is actual retinal coordinate of the horizon, where

$$y' = (y - 200)/f$$

Then the  $y'$ -coordinate of the vanishing point (and hence the horizon) can be obtained by letting  $\lambda \rightarrow \infty$ :

$$y' =_{\lambda \rightarrow \infty} \frac{Y_c - \lambda \sin \alpha}{Z_c + \lambda \cos \alpha} = \frac{-\sin \alpha}{\cos \alpha}.$$

So  $\alpha = -\text{atan}(y')$  can be directly computed from the  $y'$ -coordinate of the horizon.

b) There are couple ways how this can be done. The coordinate transformation between the feet coordinate frame  $\{w\}$  and the eye coordinate frame  $\{c\}$  has the following form

$$\mathbf{X}_w = R\mathbf{X}_c + T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha \\ 0 & -\sin \alpha & \cos \alpha \end{bmatrix} \mathbf{X}_c + \begin{bmatrix} 0 \\ -t_y \\ 0 \end{bmatrix}$$

where  $t_y$  is the height of the person and  $\alpha$  is the viewing angle. The inverse transformation is then expressed as  $\mathbf{X}_c = R^T \mathbf{X}_w - R^T T$

$$\mathbf{X}_c = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{bmatrix} \mathbf{X}_w + \begin{bmatrix} 0 \\ t_y \cos \alpha \\ t_y \sin \alpha \end{bmatrix}$$

<sup>1</sup>Horizon line is the line where all parallel lines in the road plane intersect in the image.

Suppose that the coordinate of the obstacle in the feet coordinate system is  $[X_w, 0, Z_w]^T$ , then the retinal y-coordinate  $y'$  (computed as in a) ) is related to its 3D counterpart as

$$y' = \frac{Y_c}{Z_c} = \frac{-\sin \alpha Z_w + \cos \alpha t_y}{\cos \alpha Z_w + \sin \alpha t_y}.$$

Since all the above quantities except  $t_y$  are known we can compute it as

$$t_y = \frac{y' \cos \alpha Z_w + \sin \alpha Z_w}{\cos \alpha - y' \sin \alpha}.$$

5. (15) **Rotational motion.** Consider set of corresponding points  $\mathbf{x}'_1$  and  $\mathbf{x}'_2$  in pixel coordinates in two views, which are related by pure rotation  $R$ . Assume that all the parameters of the calibration matrix  $K$  are known except the focal length  $f$ . Describe in detail following steps of an algorithm for recovering  $R$  and the focal length of the camera  $f$ .

- The projection points in two views are related by an unknown  $3 \times 3$  matrix  $H$ . Write down the parametrization of matrix  $H$  in terms of rotation matrix entries  $r_{ij}$  and the focal length  $f$ .
- Describe a linear least squares algorithm for estimation of matrix  $H$ . What is the minimal number of corresponding points needed in order to solve for  $H$  ?
- Given the parametrization of  $H$  derived in a) describe a method for estimating the actual rotation and the focal length of the camera.

### Solution

Rigid body motion equation in case of pure rotation and partial calibration can be written as

$$\lambda_2 \mathbf{x}_2 = K_f R K_f^{-1} \lambda_1 \mathbf{x}_1$$

where the intrinsic calibration matrix  $K_f$  has the following form (the only unknown is the focal length)

$$K_f = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and  $\mathbf{x}_1, \mathbf{x}_2$  are the partially calibrated image coordinates (center of the image has been subtracted and aspect ratio is assumed to be 1). Eliminating the unknown scales  $\lambda_i$  and denoting  $H = K_f R K_f^{-1}$  we obtain a following constraint between the measurements and the unknown  $H$ :

$$\widehat{\mathbf{x}}_2 H \mathbf{x}_1 = 0$$

This constraint will give us two linearly independent equations per corresponding pair which can be written in the form:

$$a_{1i} H_s = 0 \quad (4)$$

$$a_{2i} H_s = 0 \quad (5)$$

where  $H_s = [h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33}]^T$  is a stacked form of matrix  $H$ . Given at least 4 points (two equations each) we can collect all the constraints as rows of matrix  $A$  and solve for  $H_s$  as

$$A H_s = 0$$

Least squares solution to the above system of homogeneous equations is obtained as eigenvector associated with the smallest eigenvalue. Given the least squares solution of  $H_s$ , gives us  $\hat{H}$  up to an unknown scale  $\gamma$ . Now given this  $\hat{H}$  we need to recover  $K_f$  and  $R$ . Note that  $\hat{H}$  has the following form

$$H = \gamma \begin{bmatrix} r_{11} & r_{12} & f r_{13} \\ r_{21} & r_{22} & f r_{23} \\ r_{31}/f & r_{32}/f & r_{33} \end{bmatrix}$$

Given  $\hat{H}$  the focal length can be then estimated using the constraints between rows of rotation matrix as  $r_1^T r_2 = 0$ . From this constraint  $f$  can be recovered as

$$f = \sqrt{\frac{h_{11}h_{12} + h_{21}h_{22}}{-h_{31}h_{32}}}.$$

Given  $f$  we can divide the entries  $h_{13}, h_{23}$  by  $f$  and multiply the entries  $h_{31}, h_{32}$  by  $f$  to eliminate the unknown focal length factor. The unknown scale  $\gamma$  can be obtained by normalizing the rows of the new  $\hat{H}$  (corrected by dividing appropriate entries by  $f$ ) to be unit norm so as to yield a proper rotation matrix.

6. (10) **6-point Algorithm for the Recovery of Fundamental (Essential) Matrix.** The relationship between a planar homography  $H$  and the Fundamental (Essential) matrix  $F = \widehat{T'}H$  (section 5.3.4), suggests a simple alternative algorithm for the recovery of  $F$ . Outline the steps of the algorithm by assuming that you have available correspondences between at least 4 planar points and at least two points which do not lie in the plane. Proceed by first recovering  $H$ , followed by  $T'$ .

**Solution**

First recover the homography  $H$  using the 4 planar points. Since  $\mathbf{x}_2^T F \mathbf{x}_1 = 0$ , we have  $\mathbf{x}_2^T \widehat{T'} H \mathbf{x}_1 = 0$ , hence  $\mathbf{l}^T T' = 0$ , where  $\mathbf{l} = \widehat{\mathbf{x}_2} H \mathbf{x}_1$ . Therefore, one can get the epipole  $T'$  by taking an intersection of two epipolar lines  $\mathbf{l}^1$  and  $\mathbf{l}^2$  defined by the two non-planar points in the second view and their correspondences in the first view warped by homography. That is,  $T' = \widehat{\mathbf{l}^2} \mathbf{l}^1$ . Given  $H$  and  $T'$  the  $F$  becomes  $F = \widehat{T'} H$ .