Exam A - Solution

1. You are allowed to use one A4 page, written in both its sides.
2. You can use any non-graphical calculator (The use of other electronic devices such as laptops, cell phones etc. is prohibited).
3. Exam duration is 3 hours.
4. There are 4 questions worth 105 points in total, each with several sections. Please read the whole question before answering.
5. There are 7 pages in this exam.
6. Please use clear hand writing and explain every step of your answers. Only clear, well explained (correct) answers will be rewarded with full grade.

Good luck!
Question 1 - Image Descriptor (27 points)

1. Name 3 properties which the SIFT descriptor is invariant to. Shortly describe (in words) what mechanism helps achieving this invariance. (7 points)

2. For what purpose do we use Harris criterion during the SIFT Descriptor Algorithm? (4 points)

3. List the main steps of Harris corner detector algorithm. (7 points)

4. We would like to use the Harris method to detect corners in a 3D image \( I \). That is, there is an intensity value \( I(x, y, z) \) for each \((x, y, z)\) voxel (note: voxel is a 3D pixel). Describe how to generalize the Harris corner detector on a 3D image. Give the main steps of the generalized algorithm, including a test to decide when a voxel is a corner. (9 points)

Solution:

1. 
   A. Rotation
   B. Translation
   C. Scale
   D. Illumination change

2. The Harris corner detector helps to discriminate edges and corners.

3. 

ANSWER:

(a) Compute the image derivatives \( I_x \) and \( I_y \) by convolving the image \( I \) with the derivatives of a Gaussian filter \( g(x, y) \).

(b) Form the Harris matrix \( H(x, y) = \sum_{u,v} g(u,v) \begin{bmatrix} I_x^2 & I_x I_y & I_x I_z \\ I_x I_y & I_y^2 & I_y I_z \\ I_x I_z & I_y I_z & I_z^2 \end{bmatrix} \).

(c) Select points such that the Harris operator \( K(x, y) = \det(H)/\text{trace}(H) \) is above a threshold \( K(x, y) > k \).

4. 
   A. Compute the derivatives \( I_x, I_y \) and \( I_z \)
   B. Construct the Harris matrix:

   \[ H = \begin{bmatrix} I_x^2 & I_x I_y & I_x I_z \\ I_x I_y & I_y^2 & I_y I_z \\ I_x I_z & I_y I_z & I_z^2 \end{bmatrix} \]
   
   C. 
   Defining \( \alpha = \frac{\lambda_1}{\lambda_3} \), \( \beta = \frac{\lambda_2}{\lambda_3} \),

   We get 

   \[ \text{Tr}(H) = \lambda_1 + \lambda_2 + \lambda_3 \]
   \[ \text{Det}(H) = \lambda_1 \lambda_2 \lambda_3 \]
   \[ \text{Tr}(H) = \lambda_3 (\alpha + \beta + 1) \]
   \[ \text{Det}(H) = \lambda_3^3 \alpha \beta \]
(Other solutions which refers to the relation of the values of $\lambda_1, \lambda_2, \lambda_3$ got accepted)

D. For Detecting a corner we would like to find points s.t. $\alpha, \beta \approx 1$ (ant number between 0.1 to 10 will be accepted in the exam) therefore, we will detect as a corner voxels which their H matrix obeys:

$$\frac{\text{Tr}(H)^3}{\text{Det}(H)} = \frac{\lambda_1^3 (\alpha + \beta + 1)^3}{\lambda_3^3 \alpha \beta} \approx \frac{3^3}{1} = 27$$

(Other solutions which refers to the relation of the values of $\lambda_1, \lambda_2, \lambda_3$ got accepted)
**Question 2 - Epipolar Geometry** (26 points)

3 ideal cameras, denoted as cameras 1, 2, and 3, take a picture of the same scene. The cameras relative pose is known. That is, the rotation matrices $R_2, R_3$ and the translation vectors $t_2, t_3$ of cameras 2, 3 relative to camera 1, are known. Suppose the world coordinate system coincides with that of camera 1.

The 3D point $X$ is projected to the image points $x_1, x_2, x_3$ in cameras 1, 2 and 3 respectively. During all the following sections, all the points $X, x_1, x_2, x_3$ are unknown unless explicitly stated.

**Clarification:** When asking for quantitative results, use a mathematical formulation. There is no need to give an explicit solution.

a. Suppose you know $x_2$ only ($X, x_1, x_3$ are unknown). What can you say about the 3D location of $X$ (in world coordinates)? Answer both qualitatively and quantitatively. (3 points)

b. Write an explicit form for the following essential matrices:
   I. $E_{12}$ the essential matrix of cameras 1 and 2 (2 points)
   II. $E_{13}$ the essential matrix of cameras 1 and 3 (2 points)
   III. $E_{23}$ the essential matrix of cameras 2 and 3 (3 points)
   *Hint: Find the rotation matrix $R_{23}$ and translation $t_{23}$ between camera 2 and camera 3 coordinate systems.*

c. Suppose you know $x_2$ only ($X, x_1, x_3$ are unknown). What can you say about the image location of $x_3$? Answer both qualitatively and quantitatively. (5 points)

d. Suppose you know $x_1$ and $x_2$ ($X, x_3$ are unknown), what can you say about the image location of $x_3$? Answer both qualitatively and quantitatively. (5 points)

e. The epipolar constraint refers to two image points, taken from two cameras, and it must be satisfied if the two points are projections of the same 3D point. Suggest an epipolar constraint(s) for the system above which contains three image points taken from three cameras. (6 points)
Solution:

Reminders:

A point described by 3D vector $P_i$ in camera $i$ is represented by $P_j = R_{ij}(P_i - O_j) = R_{ij} P_i + T_{ij}$.

The image point in camera $j$ is (hom. Coord.) $(u,v,1)' \sim M X_i$ where the rows of $M$ are denoted $Mx, My, Mz$.

1. **The 3D point lies on a line (or a ray) specified by the pair of equations**
   
   $$(Mx)'X_1 / (Mz)'X_1 = u, \ (My)'X_1 / (Mz)'X_1 = v.$$  

2. **Knowing $x_2$ we know that $x_3$ is on the epipolar lines specified by $E_{23}$ $x_2$.**

3. **Knowing both $x_2$ and $x_1$, we know that $x_3$ is the point in the intersection of the two epipolar lines $E_{23} x_2$ and $E_{13} x_1$.**

4. **If the three points $x_1, x_2, x_3$ correspond to the same 3D point then $x_3$ should satisfy the two epipolar constraints as well as the third one: $x_2 E_{12} x_1 = 0$.**

- a. The 3D point lies on a line (or a ray) specified by the pair of equations 
  
  $$(Mx)'X_1 / (Mz)'X_1 = u, \ (My)'X_1 / (Mz)'X_1 = v.$$  

- b. 
  1. $E_{12} = (T_{12}) \times R_{12}$, 
  2. $E_{13} = (T_{13}) \times R_{13}$ 
  3. To write the third Essential matrix we need to calculate the corresponding rotation matrix and translation vector. $P_2 = R_{12} P_1 + T_{12}$ therefore $P_1 = (R_{12})' P_2 - (R_{12})' T_{12}$. $P_3 = R_{13} P_1 + T_{13} = R_{13} (R_{12})' P_2 - R_{13} (R_{12})' T_{12} + T_{13}$. Therefore 
    
    $R_{23} = R_{13} (R_{12})'$ and $T_{23} = -R_{13} (R_{12})' T_{12} + T_{13}$. 
    
    The essential matrix is $E_{23} = (T_{23}) \times R_{23}$.

- c. Knowing $x_2$ we know that $x_3$ is on the epipolar lines specified by $E_{23} x_2$.

- d. Knowing both $x_2$ and $x_1$, we know that $x_3$ is the point in the intersection of the two epipolar lines $E_{23} x_2$ and $E_{13} x_1$.

- e. If the three points $x_1, x_2, x_3$ correspond to the same 3D point then $x_3$ should satisfy the two epipolar constraints as well as the third one: $x_2 E_{12} x_1 = 0$. 


Question 3 - Recognition (27 points)

1. (6 points) Using Hough transform, how can we identify if a binary image contains parallel lines?
   Answer this question separately, for the two line parametrizations:
   i. \( y = ax + b \)
   ii. \( r = x \cos \theta + y \sin \theta \)

A Parallel Group in an image, is a group of \( n \geq 2 \) lines which are all parallel.

2. (5 points) Design an algorithm which gets a binary image and uses the Hough transform with parametrization (ii) to return the following parameters:
   i. The number of parallel group in the image
   ii. The number of parallel lines in each group
   iii. The direction of each group
   Describe in details all the algorithm steps.

3. (5 points) The Hough transform is applicable only for binary images. To apply it on gray-level images, a pre-process is needed. Describe the main steps of an algorithm which pre-process the image to enable line detection using Hough transform.

You are now given a data-set of images which contains segments of obelisks from all around the world. Some of the images are of the Technion obelisk. A sample of this data-set is shown below:

![Technion obelisk segments](image1)

![Other obelisk segments](image2)

4. (5 points) Design a classification algorithm which will label an image from this data-set as either a “Technion obelisk” or as one of the “Other obelisks”. Describe in details all the steps of your algorithm. Use your answers from the previous sections.

Assuming that:
   i. The total width of all obelisks is equal, and given by \( w \)
   ii. The width of the Technion obelisk’s bands is \( h \)
   iii. The maximal difference between the angles of the bands of the Technion obelisk is \( \delta \)
5. Describe critical parameters of the Hough algorithm and the classification algorithm, to ensure that your classification algorithm will succeed. (6 points)
Solution:

1. Assuming a binary image which have the values 1 for lines candidate:
   
      
      For each each point in the image with a value of 1, vote (+1) for all the cell which can
      represent line that passes thorough this point.
      
      Take as the lines parameters the maximal voted cells (or higher than some threshold).
      
      If, after taking the maximum, two points have the same a value- they corresponds to
      parallel lines.

   ii. Create a Hough space grid $[0:dthta:Phi,0:dr:R]$.
      
      For each each point in the image with a value of 1, vote (+1) for all the cell which can
      represent line that passes thorough this point.
      
      Take as the lines parameters the maximal voted cells (or higher than some threshold).
      
      If, after taking the maximum, two points have the same thta value- they corresponds to
      parallel lines.

2. Use the voting procedure of 1.i and create the Hough voting grid. Tae maximum
   
   Set:
   
   NUM_OF_GROUP = 0;
   NUM_OF_PARALLEL = 0;
   PARALLEL_IND = 0;
   DIRECTION = 0;

   For each theta = 0:dthta:Phi, do:
   
   Count = 0;
   For r = 0:dr:R
   
   If [theta , r] was detected as maximal point, count = count+1;
   End
   If count>1
   
   NUM_OF_GROUP = NUM_OF_GROUP+1;
   PARALLEL_IND = PARALLEL_IND+1;
   NUM_OF_PARALLEL[PARALLEL_IND] = count;
   DIRECTION[PARALLEL_IND] = theta;
   End

3. Hough Transformed operates on binary edges image. Therefore, it is necessary to use edge
   detection algorithm for example Canny:
   
   (a) Filter the image I with the derivatives of Gaussians: IGx and IGy.
   (b) Find the magnitude and the orientation of the image gradient.
   (c) Find locations in the image where the magnitude of the gradient is above a Th_high.
   (d) Thin: check neighbors in Grad direction and choose maximum
   (e) Find next: Find pixel neighbor in tangent direction
   (f) If (GradMag > Th_low) continue to Thin (4), Else continue to initialize (3)

   (also correct answers that didn’t describe all the steps of canny we re accepted)

4. Here there are a lot of correct possible solutions. The key observation is that the Technion
   obelisk contains many parallel lines in the same on direction.
   The algorithm in general is:
   
   a. transform the segments into their edges image, using section 3.
   b. perform Hough transform
   c. use the algorithm of section 2 to detect parallel groups
d. if an image contains at least one parallel group with amore then N points (N is some threshold), then it’s the Technion’s obelisk, otherwise- other obelisk.  
less points were given to solutions assume labeled data.

5. We can design the Hough grid accordingly:
   a. The theta resolution needs to be no smaller than delta, so we will have all the lines at the same group.
   b. The resolution needs to be smaller than h so we could separate all the lines.
   c. We could make the algorithm more efficient by searching for parallel group only in the range of ~W in the r direction.
Question 4 - Image Transformations (25 points)

1. For each of the images a,b,c,d:
   a. What is the type of deformation is needed to transform the source image to be aligned with it? Briefly describe what are the details within the image which indicate the deformation type. (5 points)
   b. Write the parametric form of the transformation. How many degrees of freedom this transformation has? (5 points)

2. Given two images $I_1, I_2$, and a set of $N$ matching key points $\{X^n_1, X^n_2\}_{n=1}^N$,
   a. Describe an algorithm which finds an Homography (Projective) transformation between the two images, such that $I_1$ is transformed to be aligned to $I_2$. (7 points)
      Pay attention: $X^n_1$ is a 2D point $X^n_1 = (x^n_1, y^n_1)$ which belongs to $I_1$.
      $X^n_2$ is a 2D point $X^n_2 = (x^n_2, y^n_2)$ which belongs to $I_2$.
   b. What is the minimal size of the set $N$? Why? (1 point)
   c. During the matching process, $M$ out of the $N$ pairs were matched incorrectly. Describe an algorithm to avoid these points. Give the main steps of the algorithm in details. (7 points)
Solution:

1. 

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale (mainly in y direction): the image got “squeezed” in this direction.</td>
<td>Uniform scale: The image is “zoomed out”.</td>
<td>Rotation: the picture is rotated along the image center</td>
<td>Projective (homography): parallel lines are not parallel anymore.</td>
</tr>
</tbody>
</table>
| \[
\begin{bmatrix}
S_x & 0 & 0 \\
0 & S_y & 0 \\
0 & 0 & 1
\end{bmatrix}
\] | \[
\begin{bmatrix}
S & 0 & 0 \\
0 & S & 0 \\
0 & 0 & 1
\end{bmatrix}
\] | \[
\begin{bmatrix}
\cos \theta & -\sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & 1
\end{bmatrix}
\] | \[
\begin{bmatrix}
a & b & c \\
d & e & f \\
g & h & 1
\end{bmatrix}
\] |
| (Answer with Sx=1 also got all the points) | 2 DOF                    | 1 DOF                                    | 1 DOF                                    | 8 DOF                                    |
| (or 1 if Sx=1)                             |                          |                                          |                                          |

If the (0,0) of the image is located it the left up corner, images a,b,c are also translated (+2 DOF). Answers that assumed translation got all points.

In (c) it is also possible to assume uniform scale (+1 DOF).

2. 

a. Projective transformation:

\[
\begin{bmatrix}
x'_i \\
y'_i \\
w'_i
\end{bmatrix} = 
\begin{bmatrix}
a & b & c \\
d & e & f \\
g & h & 1
\end{bmatrix} 
\begin{bmatrix}
x_i \\
y_i \\
w_i
\end{bmatrix}
\]

we will write as:

\[
\begin{bmatrix}
x'_i \\
y'_i \\
w'_i
\end{bmatrix} = 
\begin{bmatrix}
a & b & c \\
d & e & f \\
g & h & i
\end{bmatrix} 
\begin{bmatrix}
x_i \\
y_i \\
w_i
\end{bmatrix}
\]

And will normalized by i after finding all the coefficients.

\[
x'_i \rightarrow \frac{x'_i}{w'_i}
\]

\[
y'_i \rightarrow \frac{y'_i}{w'_i}
\]

Hence:

\[
x'_i = \frac{ax_i + by_i + c}{gx_i + hy_i + a}
\]

\[
y'_i = \frac{dx_i + ey_i + f}{gx_i + hy_i + a}
\]

Using the DLT form:
This is an over constrained problem, which can be solved by minimizing:

$$\arg \min_b \|Ab\|_2^2$$

s.t. $\|b\|_2 = 1$

The solution to this problem is the left eigen vector corresponds to the smallest eigen value of $A$. It can be constructed by using SVD decomposition of the matrix $A$.

After finding $b$, we will normalized it by its last entry, $I$, and reshape it to size 3x3.

b. Each pair gives 2 eq.

8DOF = 4 pairs.

c. We will use RANSAC:

Initialization:

Max_Inliers_Number=0;

Error_threshold = $t$;

For $k=1,...,K$ do:

i. Chose a subset of at least 4 pairs

ii. Calculate the transformation $H$ according to this subset.

iii. Use this transformation to warp the coordinate $X_i^n$ and calculate the distance $d(X_i^n, X_2^n)$ (by L2 norm, for example).

iv. Count the number of points $T$, which satisfy $d(X_i^n, X_2^n) \leq t$. This is the number of inliers.

v. If $T > \text{Max}_\text{Inliers}_\text{Num}$

1. $\text{Max}_\text{Inliers}_\text{Num} = T$

2. Calculate $H$ according to all inliers

Return $H$.

Solution that used the values $N$ and $M$ to set the sizes of set and stopping criteria were accepted.