Exam in SSY096

Allowed materials: Pencil, eraser.

The exam consists of six problems. Make sure that you have them all.

- Motivate all answers carefully.
- Use a new paper for each new numbered problem.
- Only write on one side of the papers.
- Write your anonymous number on each new page.
- Avoid using a red pen.

Grades

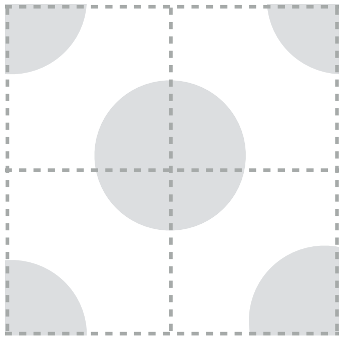
≥ 8 **points** Grade: 3

≥ 11 **points** Grade: 4

≥ 14 **points** Grade: 5

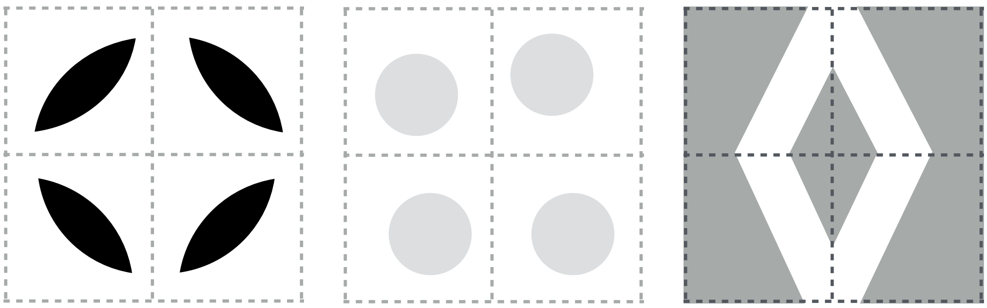
1 SIFT, 3 points

(a) A SIFT-like descriptor (as the one in Lab 1) was computed for the following image Q. The regions used are indicated with grey dashed lines so these lines are not part of the actual image. Note that unlike for original SIFT only four regions are used.

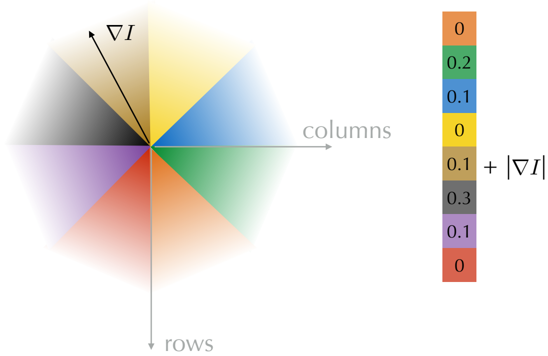


Q

Which of the following three images, A, B and C will produce the most similar SIFT-like descriptor. Motivate your answer carefully. Don't forget that a picture can help. The structure of the gradient histograms is indicated below for reference.



A B C



(b) Explain (in a few sentences) how SIFT features together with Ransac-based registration can be used to recognize buildings in images.

2 Image registration, 3 points

Let's say we want to warp a source image I_s to a target image I_t. Let $\tilde{x}_1, \ldots, \tilde{x}_n$ be points in the source image and $x_1, \ldots, x_n$ points in the target image. In order to align the images we use Ransac to estimate an *affine* transformation.

(a) Explain how to construct minimal solver for this problem. Show how to use the x_j's and $\tilde{x}_j$'s to get a system on $M\theta = b$ form, (where θ are all the unknowns) and explain how it can be solved in Matlab. Be sure to define all variables that you use in your explanation.

(b) Having estimated a solution with Ransac, it can be refined using a least-squares solver. Why do we need to run Ransac first?

(c) Shown below is a source image, I_s, a deformation field (Δ_u, Δ_v) and an empty warped image, I_w. Compute $I_w(2,3)$ in the resulting warped image. Explain how you get that value. As always u refers to rows and v to columns.

2	1	4
5	11	17
12	18	11
15	5	7
9	10	8
3	13	6

I_s

2	1	0	-1
2	1	0	-1
2	1	0	-1
2	1	0	-1
2	1	0	-1

Δ_u

-2	-2	-2	-2
-1	-1	-1	-1
0	0	0	0
-1	-1	-1	-1
-2	-2	-2	-2

Δ_v

		?	

I_w

3 Triangulation, 3 points

(a) Given the camera matrix

$$P = \begin{pmatrix} 100 & -100 & 40 & -20 \\ 100 & 100 & 60 & 20 \\ 0 & 0 & 1 & 2 \end{pmatrix}, \tag{1}$$

a 3D point

$$X = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \tag{2}$$

and a corresponding image point $(u, v) = (37, 156)$. Compute the reprojection error, that is, the length of the reprojection residual. Also check that if the depth is positive. What does it mean when a depth is negative?

(b) For a general camera matrix P, 3D point, X and image point (u, v), write down formulas for the reprojection residual. The formulas should be on the form

$$r_u = \ldots \tag{3}$$

$$r_v = \ldots \tag{4}$$

Make sure to clearly define any new variables that you introduce. Note that we want a general formula so don't use the numerical values from (a).

Assume that Ransac was used to triangulate a 3D point as in Lab 3. What would you say about the certainty of this estimate if we got

(c) 1 inlier?

(d) 2 inliers?

(e) 3 inliers?

Motivate your answers.

4 Statistical learning, 3 points

(a) When training a classifier similar to what you did in Lab 4 we want to minimize a function

$$L(\theta) = \sum_{i=1}^{n} L_i(\theta). \tag{5}$$

Assume that n is a very large number. Explain how you would do this using stochastic gradient descent. What is the big advantage compared to ordinary gradient descent? (You don't need to remember the formula for L_i. Just explain in general terms.)

(b) What happens if we have too large a learning rate, μ? What happens if the learning rate is too small? Feel free to use an illustration to explain. Is there a way to get around this dilemma?

(c) When training a classifier you normally use a training set, a validation set and a test set. Explain (briefly) the roles of the three sets. What is the risk of not using a test set?

5 Ransac, 3 points

Consider aligning two images with an affine transformation. We have extracted a set of SIFT points $x_1, \ldots, x_n$ in the target image and a set $\tilde{x}_1, \ldots \tilde{x}_n$ in the source image and reordered them such that x_i corresponds to $\tilde{x}_i$. We want to use Ransac to estimate an affine transformation. Prior experiments with the same type of images has shown that on average 87.5% of the correspondences are outliers.

(a) How many Ransac iterations, K, are required to ensure to 99% certainty, that at least once, we picked an outlier-free subset in Ransac. Answer with a formula,

$$K = \ldots, \tag{6}$$

since getting a numerical value might require a calculator.

(b) Experiments show that by using a camera with twice the resolution (four times as many pixels), the number of matched SIFT points, n, increases with a factor 4 and the rate of inliers increase from 12.5% to 25%.

How will this influence the number of iterations if we use the rule of thumb discussed in the lectures? How will it affect the total running time of Ransac? Take care to consider all steps.

6 Camera geometry, 3 points*

Consider a video camera filming a rotating object. Let

$$X(t) = R(t)X_0 = R(t)\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \tag{7}$$

be the position of a point on the object at time t. Here, $R(t)$ is a rotation $5t$ radians[1] about the axis

$$\frac{1}{5}\begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix}. \tag{8}$$

This means that the object is rotating with a constant angular velocity and in particular, that

$$R(0) = I. \tag{9}$$

The camera has camera matrix

$$P = \begin{pmatrix} 100 & -100 & 50 & 0 \\ 100 & 100 & 50 & 0 \\ 0 & 0 & 1 & 9 \end{pmatrix}. \tag{10}$$

Let $x(t)$ be the 2D-projection of $X(t)$ in this camera, and compute the 2D velocity vector, $x'(t)$, at time $t = 0$, that is, compute how the projection moves in the image.

[1] You can choose the rotation direction.