MTU
Ollscoil Teicneolaíochta na Mumhan
Munster Technological University

MACHINE VISION

Assignment 1:
Feature points and descriptors

DUE DATE

This assignment should be submitted to Canvas before 11:59pm on **Friday 19/03/2021**.

Please submit a single ZIP file with your student number and name in the filename. Your submission should contain **exactly 2 files**:

- A detailed documentation of all code you developed, including the tests and evaluations you carried out. Please make sure that you include a .pdf document with every result image you produce referencing the exact subtask and lines of code it refers to.
- All Python code you developed in a single .py file that can be executed and that generates the outputs you are referring to in your evaluation. The file needs to be readable in a plain text editor, please do NOT submit a notebook file or link. Please also make sure that you clearly indicate in your comments the exact subtask every piece of code is referring to.

Please do NOT include the input files in your submission.

You can achieve a total of 50 points as indicated in the tasks.

TASK 1 (Scale space, 11 points)

In this assignment you will implement a scale and rotation invariant point feature extraction algorithm inspired by the scale invariant feature transformation algorithm (SIFT) to identify a set of interest points in an image together with their respective scales and rotations.

A. Download the input image file **Assignment_MV_1_image.png** from Canvas. Load the file and convert it into a single channel grey value image. Make sure the data type is float32 to avoid any rounding errors. [2 points]

B. Create twelve Gaussian smoothing kernels with increasing scales

$$\sigma \in \{2^{k/2} \mid k = 0, \dots, 11\}$$

and plot each of these kernels as image [7 points]. Make sure that the window size is large enough to sufficiently capture the characteristic of the Gaussian. Apply these kernels to the input image to create a scale-space representation. Display all resulting scale-space images [2 points].

TASK 2 (Feature point locations, 10 points)

A. Use the scale-space representation from task 1 to calculate difference of Gaussian images at all scales where this is possible. Display all resulting DoG images [5 points].
B. Find key-points by thresholding all DoGs from subtask A. Use a threshold of $T = 10$ and suppress non-maxima in scale-space by making sure that the key-points have no neighbours, both in space as well as in scale, where the value is higher [3 points]. The resulting key-points should comprise three coordinates (x, y, σ), two spatial and the scale at which they were detected.
C. Visualise the key-point locations and their scales in the input image by drawing a circle of radius 3σ around every key-point [2 points].

TASK 3 (Feature point orientations, 17 points)

A. Calculate derivatives of all scale-space images from task 1 using the kernels

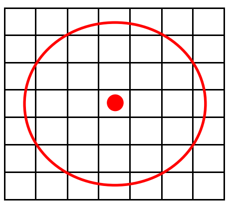
$$d_x = (1 \quad 0 \quad -1)$$

and

$$d_y = (1 \quad 0 \quad -1)^T$$

Display the resulting derivative images g_x and g_y at all scales [3 points].

B. For each key-point (x, y) consider the 7×7 grid of points sampled at a distance of $\pm\frac{9}{2}\sigma$ around its location (see figure left)



$$(q, r) \in \left\{\frac{3}{2}k\sigma \mid k = -3, \dots, 3\right\} \times \left\{\frac{3}{2}k\sigma \mid k = -3, \dots, 3\right\}$$

and calculate the gradient lengths

$$m_{qr} = \sqrt{g_x^2[x+q, y+r] + g_y^2[x+q, y+r]}$$

and gradient directions

$$\theta_{qr} = atan2\left[g_y[x+q, y+r], g_x[x+q, y+r]\right]$$

for each point on this grid [6 points]. Make sure to use the appropriate scale σ and the correct gradient images g_x and g_y. Use nearest neighbour interpolation to sample the gradient grid.

C. Calculate a Gaussian weighting function

$$\mathrm{w_{qr}} = \mathrm{e}^{-(\mathrm{q}^2+\mathrm{r}^2)/(9\sigma^2/2)}/(9\pi\sigma^2/2)$$

for each of the 7 × 7 grid points [2 point]. Then create a 36-bin orientation histogram vector h and accumulate the weighted gradient lengths $\mathrm{w_{qr}m_{qr}}$ for each grid point (q,r) where the gradient direction θ_{qr} falls into this particular bin [3 points], i.e. calculate for each 10^o range $-18 \le i < 18$ of potential direction the weighted contributions that fall into this range

$$h_i = \sum_{i \le \frac{36\,\theta_{qr}}{2\pi} < i+1} \mathrm{w_{qr}m_{qr}}$$

Use the maximum of this orientation histogram h to determine the dominant orientation

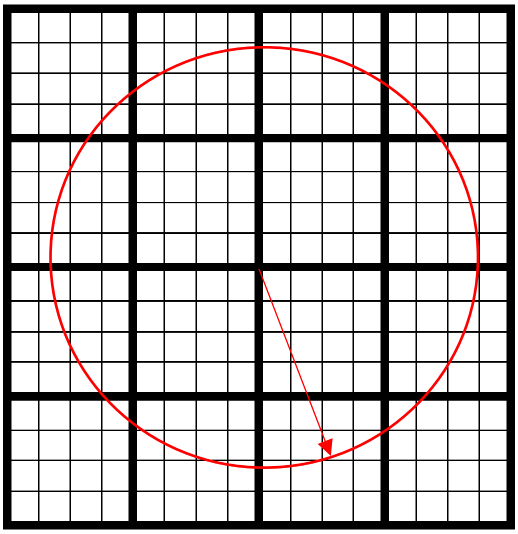
$$\hat{\theta} = \frac{2\,\pi}{36}\left(\frac{1}{2} + argmax_i\; h_i\right)$$

of the key-point [1 point]. Each key-point is now characterised by its location, its scale, and its dominant orientation $(x, y, \sigma, \hat{\theta})$.

D. Visualise the orientation of all key-points by drawing a circle with radius 3σ and a line from the key-point centre to the circle radius, which indicate the orientation (see example visualisation for a single key-point on the right). Display the resulting output image with all the key-points [2 points].

TASK 4 (Feature descriptors, 12 points)

The SIFT algorithm also proposed to calculate a 128-vector that describes the local distribution of gradient directions relative to the dominant direction for each key point. This descriptor can be used for feature matching or as scale and rotation invariant image descriptor. Similar to the orientation calculation we look at a 16 × 16 grid around each key-point location (x, y) now, again taking the scale into consideration. The 16 × 16 grid is sub-divided into 4 × 4 grids of size 4 × 4 each (see figure on the right).



A. The grid coordinates relative to the key-point location for each of the 4×4 sub-grids $(i,j) \in \{-2,\ldots,1\} \times \{-2,\ldots,1\}$ covering an area of $\pm\frac{9}{2}\sigma$ are given by

$$(s,t)_{ij} \in \left\{\frac{9}{16}\left(k+\frac{1}{2}\right)\sigma \mid k = 4i,..,4i+3\right\} \times \left\{\frac{9}{16}\left(k+\frac{1}{2}\right)\sigma \mid k = 4j,..,4j+3\right\}$$

   Calculate for each of these coordinates the Gaussian weighting function

$$w_{st} = e^{-(s^2+t^2)/(81\sigma^2/2)}/(81\pi\sigma^2/2)$$

   as well as the gradient lengths

$$\mathrm{m_{st}} = \sqrt{g_x^2[x+s,y+t] + g_y^2[x+s,y+t]}$$

   and gradient directions

$$\theta_{\mathrm{st}} = atan2\left[g_y[x+s,y+t], g_x[x+s,y+t]\right] - \hat{\theta}$$

   adjusted by the dominant direction $\hat{\theta}$ around each key-point [8 points]. Make sure to use the appropriate scale σ and the correct gradient images g_x and g_y. Use nearest neighbour interpolation to sample the gradient grid.

B. Now create a 8-bin orientation histogram vector $\mathrm{h_{ij}}$ for each of the 4×4 sub-grids $(i,j) \in \{-2,\ldots,1\} \times \{-2,\ldots,1\}$ and accumulate the weighted gradient lengths $\mathrm{w_{st}m_{st}}$ for each grid point (s,t) within the sub-grid where the adjusted gradient direction $\theta_{\mathrm{st}} - \hat{\theta}$ falls into this particular bin [3 points].

C. Concatenate all these 16 histogram 8-vectors into a single 128-vector d describing the feature at the key-point. Normalise this descriptor vector dividing it by its length $\frac{d}{\sqrt{d^T d}}$ and compute the final descriptor vector for each key-point by capping the maximum value of the vector at 0.2 [1 point].