Computer Vision 1: Homework 10

Deadline 17.01. 12:15

Important: Submit your programming solutions through Moodle. The deadline for submitting your work is always on Thursday, at 12.15, the week after handing out the homework.
For other, non-programming homework, bring your solution with you to the exercise class. For each homework problem, one student will be chosen at random to present their solution.

Programming tasks.

- From the last homework, we have three different images of the Hamburg Elbphilharmonie `im, im1, im2`. Using `skimage.feature.hog`, generate three HOG feature vectors `v, v1, v2` for each of these images, respectively. The parameters are `orientations = 8, pixels_per_cell=(16, 16), cells_per_block=(1, 1)`.
- Why do the feature vectors have size 9600?
- Compute the Euclidian distances from `v` to `v1` and `v2`. Which one is smaller?
- Generate a new augmented image `im3` by adjusting the brightness of the image. We make the image darker by using a technique called `Gamma Correction` with $\gamma = 2$. Using `skimage.exposure.adjust_gamma`, compute and visualize this new image.
- Extract HOG feature vector of `im3` as `v3` (same parameters). What is the distance between `v` and `v3`? How does it compare to the distances between `v` and the feature vectors of the affine transformed images?

  Note: write your answers to the non-programming questions above into a block of comments in your submission at either the end of the file or after the code block of the corresponding programming tasks.

Other tasks.

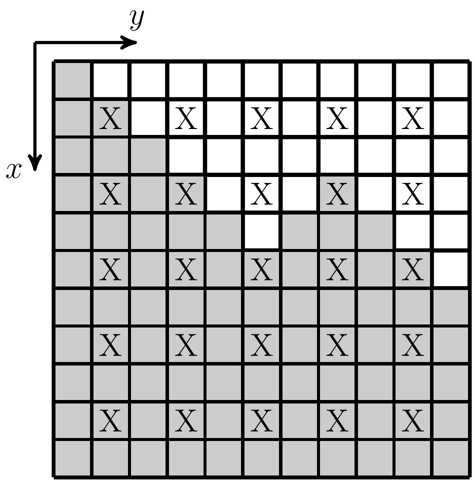
1. Keypoint orientation in SIFT is used to represent a keypoint relative to its orientation thus achieving invariance to image rotation. The orientation is calculated by using sample points around the keypoint. Given a Gaussian smoothed image L, the gradient $\nabla f(x, y) = \begin{bmatrix} g_x & g_y \end{bmatrix}^T$ at pixel location (x, y) is calculated using the central difference filter:

$$g_x = L(x+1, y) - L(x-1, y)$$
$$g_y = L(x, y+1) - L(x, y-1).$$

   Consider the image shown below, where white pixels have the value 1 and pixels with gray background have value 0. The axes labeled with x and y indicate the directions in which the respective pixel coordinates increase, so that the upper left pixel is at coordinate $(0, 0)$, and the lower right pixel is at coordinate $(10, 10)$.

   For each sample point marked by a symbol "X" in the picture, calculate

   - the gradient $\nabla f(x, y) = \begin{bmatrix} g_x & g_y \end{bmatrix}^T$,
   - the magnitude of the gradient $m(x, y) = \sqrt{g_x^2 + g_y^2}$, and
   - the angle of the gradient $\theta(x, y) = \arctan\left(\frac{g_y}{g_x}\right)$.



2. Consider Figure 1, which shows a normalized orientation histogram for a SIFT keypoint after weighting[1].

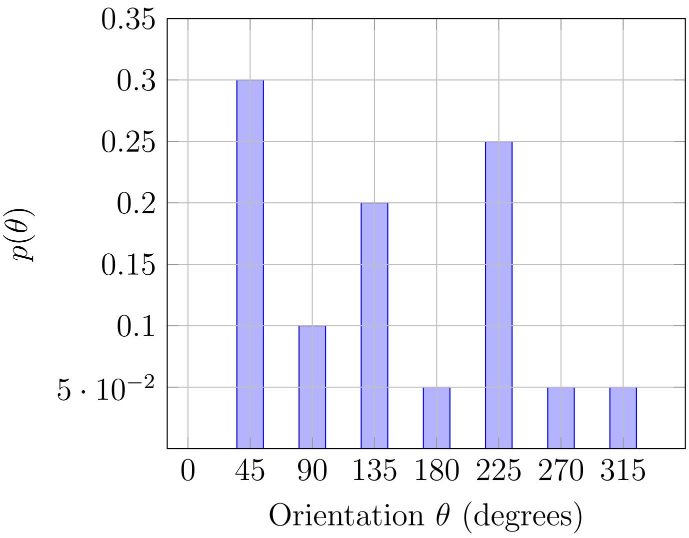


Figure 1: A normalized orientation histogram of a SIFT keypoint.

   (a) What is the dominant local direction of the keypoint?

   (b) How many new keypoints will be created, and why? What are their orientations?

3. A detector finds $m_1 = 113$ features in the first image I_1, and $m_2 = 88$ features in the second image I_2. There is a total of $n = 53$ features in I_1 that have a correspondence in I_2. There are $k = 21$ true positive matches, and $l = 32$ negative false matches.

   (a) Calculate the recall and precision.

   (b) Calculate the repeatability $R(I_1, I_2)$.

[1]For simplicity, we consider an 8-bin orientation histogram. In the original SIFT algorithm, 36 bins are used.