Nicolas Pfitzer 21 nov. 2023

Lab 5 – Computer Vision: Image Segmentation

1. Mean-Shift Algorithm:

A- Procedure

The mean-shift algorithm is an unsupervised learning algorithm most widely used for clustering and image segmentation and can be broken down into three main steps:

a) **Distance function**: The distance function computes the distance (intensity-wise) from a given pixel to all other pixels. The function returns a list of N-distances.
b) **Defining the kernel function:** Each pixel is represented by a kernel function (gaussian in this case) centered at that point, which assigns a weight to all the other pixels in the image. The gaussian has a predefined bandwidth (standard deviation) and is a function of the distance to the center pixel. Thus, pixels that closely resemble the center pixel will have higher weight in the calculation of the mean.
c) **Shifting the means:** For each pixel, we calculate the weighted mean within the kernel and shift the new pixel value to the mean.

$$Next\ pixel\ value: f(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x) x_i}{\sum_{x_i \in N(x)} K(x_i - x)}$$, where K is the kernel function (gaussian)

By repeating this process iteratively, the algorithm continues shifting the data points until convergence is reached. At convergence, the mean shift vectors become very small or negligible.

B- Results and effect of bandwidth variations:

a) Original image and default BW results

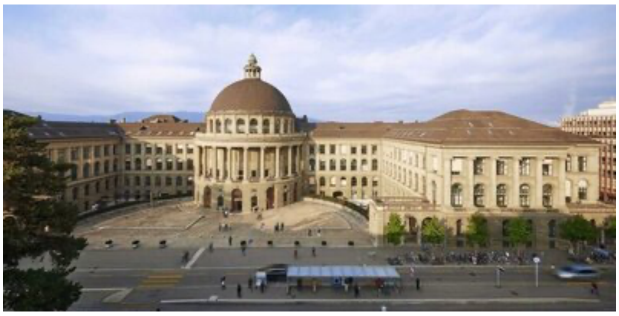

Figure 2 Original Image (ETHZ HG)

Figure 1 Mean-shift with 20 iterations and BW=2.5

To save on computation time, the scale of the image was reduced further to 0.3.

b) Bandwidth

In an ideal case, we would want the mean-shift algorithm to segment all the features in the scene. However, seeing that the resolution of the image is quite low, the aim was to focus primarily on its 5 main features:

- The sky
- The building (Walls + Roof)

- The road
- The trees

With that in mind, various combinations of bandwidth where attempted to get as close as possible to the objective.

Figure 3 BW=1 	*Figure 4 BW =3* 
Figure 5 BW =5 	*Figure 6 BW =7* 

From the results above, the best performing bandwidth seems to be BW=5 (Fig.5), as we can still clearly distinguish the key features of the image, with minimal noise.

The most important drawback from these results is that the size of the kernel (ie. Bandwidth) directly affects the precision of the image segmentation. A higher bandwidth will capture less features but will be less influenced by noise when compared to a shorter bandwidth. Therefore, depending on the number of features we want to be clearly segmented in the image and if the resolution of the image allows, we might want to reduce the bandwidth to capture finer details.

2. SEG-NET

Model accuracy (IoU score): 0.870