Computer Vision

Exercise 5: Image Segmentation

Viviane Yang 16-944-530

7th November 2019

1 Image Processing

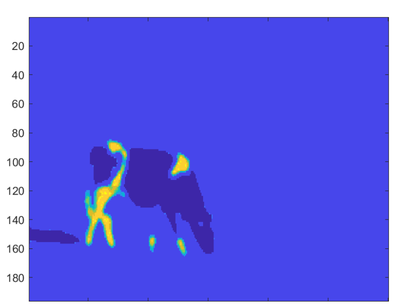
1.1 Smoothing and Conversion to L*a*b - space

The picture needs to be smoothed in order to filter out any noise or irregularities that could create too many clusters in mean shift. By using a 5×5 Gaussian matrix with $\sigma = 5.0$, the image is made more coarse grained with is better for grouping parts of the image together. As a second step, the RGB image is converted to L*a*b values. In L*a*b space, there is the L-axis for brightness and on the a-axis the complementary colors green and red and on the b-axis blue and yellow. This way, pixels of the same color family but with different brightness are closer together than in RGB space, which makes sense for segmentation.

Figure 1: Smoothed by a Gaussian filter $\sigma = 5.0$ and transformed into L*a*b space

2 Mean-Shift Segmentation

The mean-shift algorithm consists of two parts, finding the peak of each pixel and merging the peaks together if they are close enough together, which means their Euclidean distance is smaller than $r/2$, where r is the radius of spherical window. By testing out a few values, $r = 30$ seems to work fine with $K = 4$ clusters.



(a) $r = 20, K = 10$

(b) $r = 30, K = 4$

(c) $r = 40, K = 1$

The smaller r, the finer the mean-shift algorithm works, thus more clusters are created. However, we do not want to have to much clusters, thus r can be increased to a point, where we have sufficient clusters and not every point is grouped to one as with $r = 40$. The peak finding algorithm terminates if the shift made in a iteration is smaller or equal to $threshold = 1$, which means it converges if the peak found in the next iteration has a smaller distance to the last one by one pixel and need to be carried out for each pixel on the image plan. This is quite costly, which is why the algorithm runs for 3 min.

3 EM Segmentation

The third part of this exercise is to implement the Expectation Maximization algorithm, which uses a mixture Gaussian model to cluster parts of the image together. Unlike mean-shift, EM needs to know the number of clusters K. I will run the EM algorithm with the values $K = 3, 4, 5$. In the expectation step, we calculate at each pixel the probability of seeing x_l or the L*a*b values under the assumption of knowing the parameters $\Theta = (\alpha, \mu, \Sigma)$. In the maximization step, we estimate the new Θ that is most likely with the estimations from the expectation step. By alternating the two steps, we are converging to a better fit. We stop if the difference in the shift of the mean is smaller than a threshold. Here are the values for Θ for $K = 3, 4, 5$ and a threshold of 1:

3.1 K=3

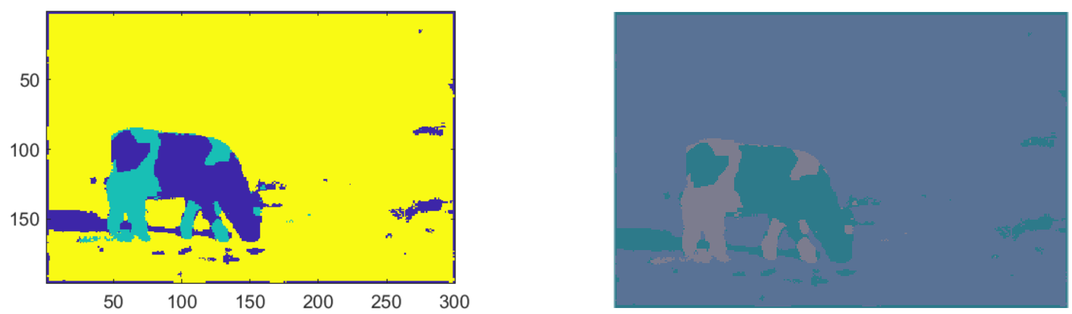


Figure 3: EM with $K = 3, thres = 1, iter = 13$

$$\Sigma_1 = \begin{pmatrix} 661.3474 & -138.0350 & 200.9529 \\ -138.0350 & 37.4786 & -48.3736 \\ 200.9529 & -48.3736 & 69.0003 \end{pmatrix}$$

$$\mu = \begin{pmatrix} 45.2556 & 121.7779 & 138.4968 \\ 124.2440 & 123.7247 & 142.1671 \\ 89.3684 & 114.4542 & 149.1027 \end{pmatrix}$$

$$\Sigma_2 = 10^3 \cdot \begin{pmatrix} 2.4268 & 0.1168 & -0.0301 \\ 0.1168 & 0.0180 & -0.0154 \\ -0.0301 & -0.0154 & 0.0324 \end{pmatrix}$$

$$\alpha = \begin{pmatrix} 0.1361 & 0.0493 & 0.8146 \end{pmatrix}$$

$$\Sigma_3 = \begin{pmatrix} 53.5574 & -0.0434 & 0.5231 \\ -0.0434 & 0.8008 & -0.1566 \\ 0.5231 & -0.1566 & 1.5483 \end{pmatrix}$$

With only 3 clusters, part of the ground is put in the same class as the dark part of the cow.

3.2 K=4

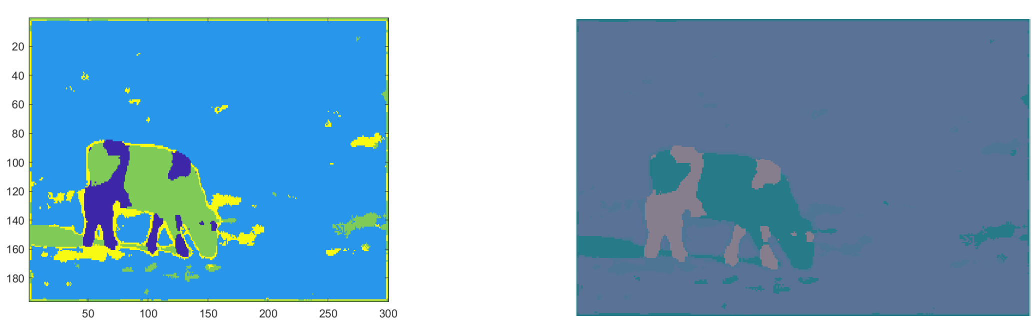


Figure 4: EM with $K = 4, thres = 1, iter = 17$

$$\Sigma_1 = 10^3 \cdot \begin{pmatrix} 2.7911 & 0.0282 & 0.0923 \\ 0.0282 & 0.0087 & -0.0063 \\ 0.0923 & -0.0063 & 0.0252 \end{pmatrix}$$

$$\mu = \begin{pmatrix} 133.5600 & 125.9130 & 139.8210 \\ 89.7125 & 114.4090 & 149.0965 \\ 38.9665 & 122.7203 & 136.8543 \\ 78.4256 & 116.6831 & 148.0148 \end{pmatrix} \qquad \Sigma_2 = \begin{pmatrix} 49.7177 & 0.2977 & 0.5426 \\ 0.2977 & 0.7327 & -0.1724 \\ 0.5426 & -0.1724 & 1.4941 \end{pmatrix}$$

$$\alpha = \begin{pmatrix} 0.0383 & 0.7911 & 0.1110 & 0.0595 \end{pmatrix} \qquad \Sigma_3 = \begin{pmatrix} 560.6493 & -130.1035 & 180.7993 \\ -130.1035 & 37.8379 & -47.5232 \\ 180.7993 & -47.5232 & 65.1307 \end{pmatrix}$$

$$\Sigma_4 = \begin{pmatrix} 91.8561 & 6.4266 & 7.3912 \\ 6.4266 & 2.5353 & -1.6007 \\ 7.3912 & -1.6007 & 8.6981 \end{pmatrix}$$

There is still a green part on the right which is clearly grass but gets classified as the same class as the cow's. However, the grass part at the cow's hind legs are now classified correctly.

3.3 K=5

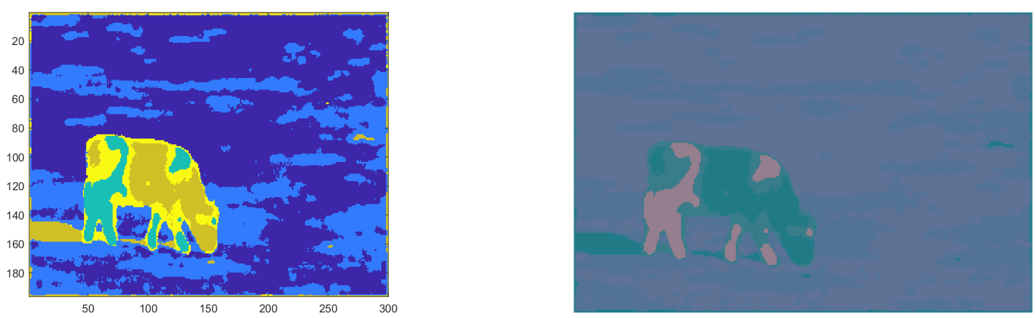


Figure 5: EM with $K = 5, thres = 1, iter = 52$

$$\Sigma_1 = \begin{pmatrix} 21.0527 & -0.1285 & 0.9765 \\ -0.1285 & 0.4848 & -0.1738 \\ 0.9765 & -0.1738 & 1.2954 \end{pmatrix}$$

$$\Sigma_2 = \begin{pmatrix} 50.9616 & -0.1089 & 5.0490 \\ -0.1089 & 1.9553 & -0.2474 \\ 5.0490 & -0.2474 & 2.3232 \end{pmatrix}$$

$$\mu = \begin{pmatrix} 92.8851 & 114.4511 & 148.9287 \\ 81.5344 & 114.5298 & 149.3815 \\ 149.2998 & 125.9690 & 140.5227 \\ 32.8776 & 123.2103 & 135.4618 \\ 56.9669 & 123.5886 & 138.3516 \end{pmatrix} \qquad \Sigma_3 = 10^3 \cdot \begin{pmatrix} 2.0300 & 0.0432 & 0.0411 \\ 0.0432 & 0.0075 & -0.0051 \\ 0.0411 & -0.0051 & 0.0249 \end{pmatrix}$$

$$\alpha = \begin{pmatrix} 0.5441 & 0.3041 & 0.0310 & 0.0731 & 0.0477 \end{pmatrix} \qquad \Sigma_4 = \begin{pmatrix} 422.1473 & -96.9735 & 143.4886 \\ -96.9735 & 25.8356 & -36.6091 \\ 143.4886 & -36.6091 & 54.3419 \end{pmatrix}$$

$$\Sigma_5 = \begin{pmatrix} 782.4852 & -129.8245 & 172.5425 \\ -129.8245 & 31.5749 & -36.4654 \\ 172.5425 & -36.4654 & 49.1237 \end{pmatrix}$$

There is now a very clear contour around the cow and the misclassified part of the grass on the right part of the picture has shrunken by a lot. The three colours light and dark yellow and turquoise blue are part of the cow. And clearly the royal blue and light blue clusters are grass. Additionally the larger K, the more iterations are needed for convergence at the same threshold.

3.4 Zebra

With EM, since we are using probability distributions and not the a local criteria such as in mean shift, the algorithm also performs well on pictures with clusters that have other clusters in-between, such as a zebra pattern.

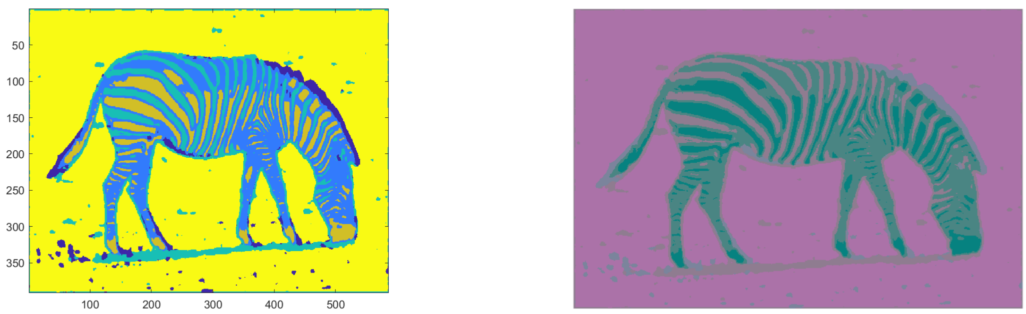


Figure 6: EM with $K = 5, thres = 2, iter = 15$

By choosing a threshold of 2 and the number of clusters $K = 5$, We can successfully segment the image of the zebra.

$$\mu = \begin{pmatrix} 122.6975 & 131.7999 & 156.1663 \\ 73.6379 & 132.9685 & 131.3126 \\ 143.2899 & 124.0617 & 144.3069 \\ 5.9030 & 129.8539 & 128.4302 \\ 170.5040 & 113.7636 & 168.4686 \end{pmatrix}$$

$$\alpha = \begin{pmatrix} 122.6975 & 131.7999 & 156.1663 \\ 73.6379 & 132.9685 & 131.3126 \\ 143.2899 & 124.0617 & 144.3069 \\ 5.9030 & 129.8539 & 128.4302 \\ 170.5040 & 113.7636 & 168.4686 \end{pmatrix}$$

$$\Sigma_1 = 10^3 \cdot \begin{pmatrix} 2.3363 & -0.1677 & 0.3353 \\ -0.1677 & 0.0623 & -0.0264 \\ 0.3353 & -0.0264 & 0.0750 \end{pmatrix}$$

$$\Sigma_2 = 10^3 \cdot \begin{pmatrix} 1.8308 & -0.0067 & 0.0035 \\ -0.0067 & 0.0259 & 0.0010 \\ 0.0035 & 0.0010 & 0.0165 \end{pmatrix}$$

$$\Sigma_3 = 10^3 \cdot \begin{pmatrix} 2.1803 & 0.1297 & -0.3179 \\ 0.1297 & 0.0435 & -0.0783 \\ -0.3179 & -0.0783 & 0.1941 \end{pmatrix}$$

$$\Sigma_4 = \begin{pmatrix} 34.7956 & 5.3813 & 4.9136 \\ 5.3813 & 3.0727 & 0.9108 \\ 4.9136 & 0.9108 & 1.9435 \end{pmatrix}$$

$$\Sigma_5 = \begin{pmatrix} 137.0011 & 12.2574 & -13.9353 \\ 12.2574 & 20.4802 & -10.7913 \\ -13.9353 & -10.7913 & 17.0035 \end{pmatrix}$$