# One-dimensional MeanShift Algorithm

Da Xu

Explore the clustering rules of the meanshift algorithm under different bandwidths.

October 8, 2023

Experiment procedure

- ▶ In order explore the clustering rules, I first apply MeanShift algorithm to the lena picture under different bandwidths. Then label the result and use heat maps to more intuitively display clustering results.

- ▶ Through experiments, I found that the main factors affecting the results are: Bandwidth, Data Consistency.

- ▶ Therefore, I smooth data using cubic spline interpolation, and use the control variable method to continuously adjust the interpolation and bandwidth and observe the results. What's more, to verify the reliability of the results, I also apply it on mandrill picture.
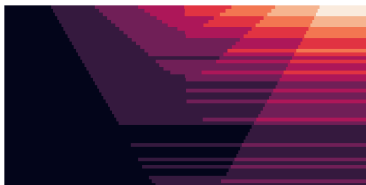
Experiment result on lena
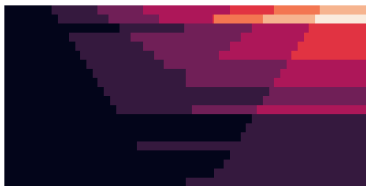
(a) Data 256 bandwidth 20
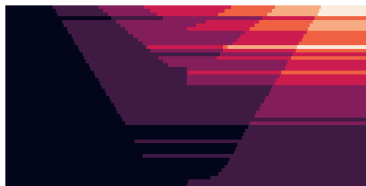
(b) Data 256 Bandwidth 50

(c) Data 1000 Bandwidth 50

(d) Data 1000 Bandwidth 100

Figure: Heat map of clustering results under bandwidth $6 \sim 20$

Experiment analysis

观察上图可以发现，当数据样本较少时及带宽精度较低时，由于数据过于离散，导致热力图出现明显锯齿。在通过三次样条插值连接平滑数据，并增加带宽精度后，可以从图中大致观察到聚类结果的变化趋势，为了更好的分析结果，将原来的灰度直方图与聚类结果热力图合并，得到下图。
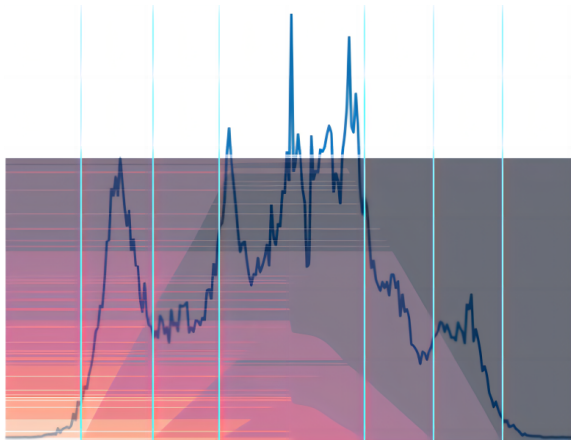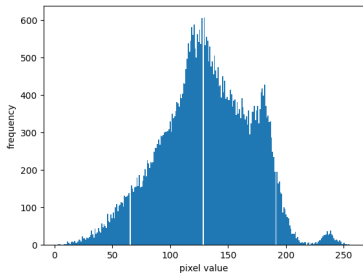
Experiment result
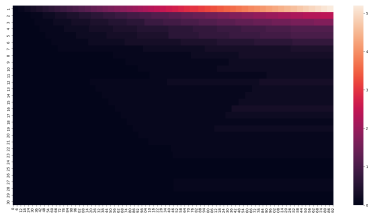

Figure: Clustering result analysis

Experiment analysis

观察上图可以发现，在带宽逐渐增大时，聚类数目有规律的减少，近似于高斯核函数曲线。另外，整体呈对称状，虽然聚类变化趋势较为明显，但仍然观察到许多突变值。在实验过程中，测试了许多方法均未解决，推测可能原因为数据不够光滑。为了验证该结论，后面将meanshift算法应用到mandrill图片上。
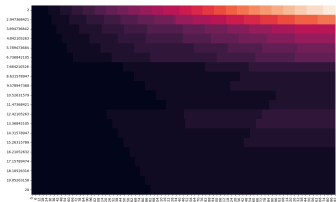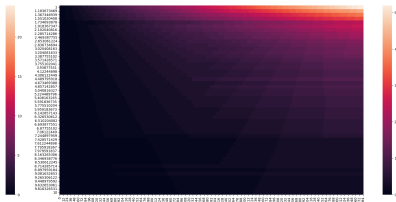
Experiment result on mandrill

(a) Original histogram

(b) bandwidth $1 \sim 30$

(c) bandwidth $2 \sim 20$

(d) bandwidth $1 \sim 10$

Figure: Heat map of clustering results