Report for Computer GraphicII, HW2
The Application of K-means Clustering

Jiang Xinhong 2020533075

October 24, 2022

Acknowledgements:

Deadline: 2022-10-25 23:59:59

You should answer the questions in **English**

You can choose C++ or Python, and no restrictions on programming framework. You can freely use frameworks such as openGL.

The **report** submits as a PDF file to gradscope, the programming part should package all the files include code, input files, executable file, readme.txt, and report. The **package** name is **your_student_name+student_id.zip**.

You will get Zero if the code not passing the plagiarism check.

1 The Application of K-means Clustering

Explore the application of K-means clustering in image superpixel segmentation and mesh simplification. **Give details of the algorithms, experimental results, and analysis**.

1.1 image superpixel segmentation (50 points)

1.1.1 Description of naive K-means Clustering algorithm

Given a set of observations (x1, x2, ..., xn), where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k ($\leq$n) sets $\mathbf{S} = \{S_1, S_2, ..., S_k\}$ so as to minimize the within-cluster sum of squares (WCSS) (i.e. variance). Formally, the objective is to find:

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \arg\min_{\mathbf{S}} \sum_{i=1}^{k} |S_i| \operatorname{Var} S_i$$

where μ_i is the mean of points in S_i. This is equivalent to minimizing the pairwise squared deviations of points in the same cluster:

$$|S_i| \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \sum_{\mathbf{x} \neq \mathbf{y} \in S_i} \|\mathbf{x} - \mathbf{y}\|^2$$

The equivalence can be deduced from identity $|S_i| \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \sum_{\mathbf{x} \neq \mathbf{y} \in S_i} \|\mathbf{x} - \mathbf{y}\|^2$.
Since the total variance is constant, this is equivalent to maximizing the sum of squared deviations between points in different clusters (between-cluster sum of squares, BCSS), This deterministic relationship is also related to the law of total variance in probability theory.

1.1.2 Algorithm steps of K-means

The algorithm steps of K-means are:

1. Select the initial k samples as the initial cluster center;
2. For each sample in the dataset, calculate its distance to k cluster centers and classify it into the class corresponding to the cluster center with the smallest distance;
3. For each class, recalculate its cluster center (ie, the centroid of all samples belonging to that class);
4. Repeat steps 2 and 3 above until a certain termination condition (number of iterations, minimum error change, etc.) is reached.

For which the pseudo-code is:

```
while cluster changed do
    for all data dots do
        for all centroids do
        end
        if p1p2 across with 2 edge then
            interpoly[i] = True /* corresponding Bool value          */
        end
    end
end
```

1.2 mesh simplification (50 points)