Homework Four Report

Zifan Wang

March 20, 2019

1 Problem 1: Texture Analysis

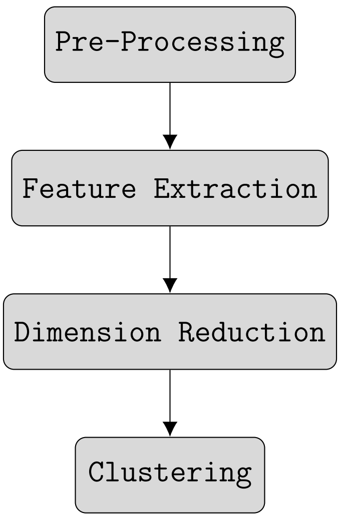
1.1 A. Texture Classification

1.1.1 Abstract and Motivation

Humans' recognition system uses two features to analysis and recognize object. The first is contour which is objects' shape and another is texture of appearances. People develop the same idea to help computer recognize objects and classify them. Texture classification refers to categorize unknown sample image into one of a set of known texture image. Nowadays, there are two methods for texture classification problem: filter and machine learning based and deep learning based methods. In this experiment, the filter and machine learning based method is developed.

1.1.2 Approach and Procedures

Texture classification process has the learning phase and the recognition phase. The learning phase is how to use the data set to help classifier categorize different textures. The following flow chart represents processes in the learning stage.



In the pre-processing stage, read each of the training images and subtract the mean pixel value from it. Subtracting the mean pixel value is to reduce pixels with high energy but less information. Apply tensor product of 5 1D kernels for 5×5 Laws filters to generate 25 5×5 Laws filter banks. The table 1 shows the five 1D kernels of Laws filters [1]:

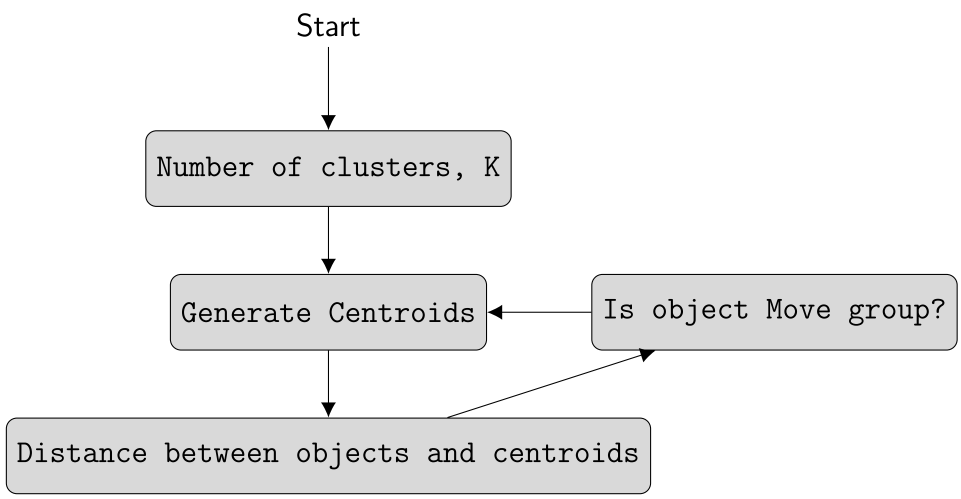
Name	Kernel
L5(level)	[1 4 6 4 1]
E5(Edge)	[-1 -2 0 2 1]
S5(Spot)	[-1 0 2 0 -1]
W5(Wave)	[-1 2 0 -2 1]
R5(Ripple)	[1 -4 6 -4 1]

Table 1: 1D Kernel for 5x5 Laws Filters

In the feature extraction stage, each training image is extended it boundary and convolved with 25 laws filters. This will return 25 output images corresponding to 25 laws filters for each training image. Then, use the following equation to calculate energy response of each filter:

$$E = \frac{1}{H \cdot W} \sum_{i=0}^{H} \sum_{j=0}^{W} (I(i,j))^2$$

where H is image height and W is image width. After calculating the energy response, a twenty five dimensional vector will be generated for each training image and we use generated energy response vectors with normalization to cluster textures.
In the dimension reduction stage, the principal component analysis (PCA) is performed to reduce less significant responses in the 25D vectors. The reason is the high dimensional feature vectors usually contains lots of redundancy. The responses with high eigen-value are enough to represent this type of textures. In addition, dimension reduction can help to visualization. Apply OpenCV library's PCA function to calculate the principal components and project 25D vector onto a 3D vector.
In the clustering stage, the kmean method is applied to cluster each texture. Use random initization to create initial centroids. The following flow chart shows the procedures of kmean clustering.



The OpenCV library kmean clustering method is also used in this experiment to group different objects.

1.1.3 Result

Texture	Actual Label	Attempt 1	Attempt 2	Attempt 3	Attempt 4
1	bubbles	1	1	1	1
2	straw	1	0	1	0
3	brick	2	3	2	3
4	bark	1	2	1	2
5	bubbles	1	1	1	1
6	bark	1	2	1	2
7	bubbles	1	1	1	1
8	straw	1	0	1	0
9	brick	3	3	3	3
10	straw	3	3	3	0
11	brick	0	3	0	3
12	bark	1	0	1	2

Table 2: Texture Classification Result

where attempt 1 uses 25D response vectors and 10 iterations, attempt 2 uses 3D response vectors and 10 iteration, attempt 3 uses 25D response vectors and 40 iterations, and attempt 4 uses 3D response vectors and 10 iteration.

Attempt	bubbles	straw	brick	bark
1	100%	0%	33.3%	0%
2	100%	66.67%	100%	66.67%
3	100%	0%	33.3%	0%
4	100%	100%	100%	100%

Table 3: Texture Classifier Performance

The figure 3 shows the performance of texture classifier. The PCA method will help the texture classification.

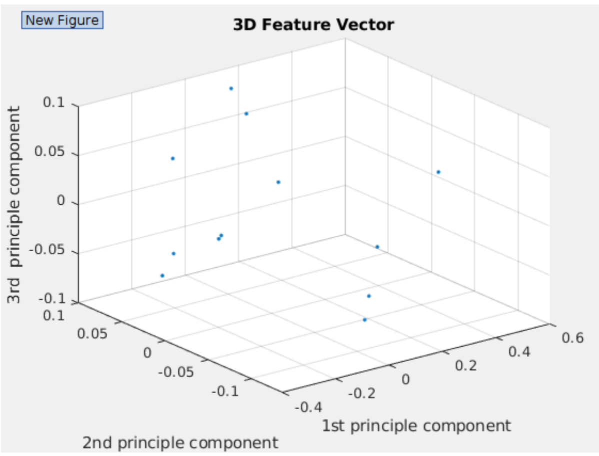


Figure 1: 3D Feature Vector Visualization

1.1.4 Discussion

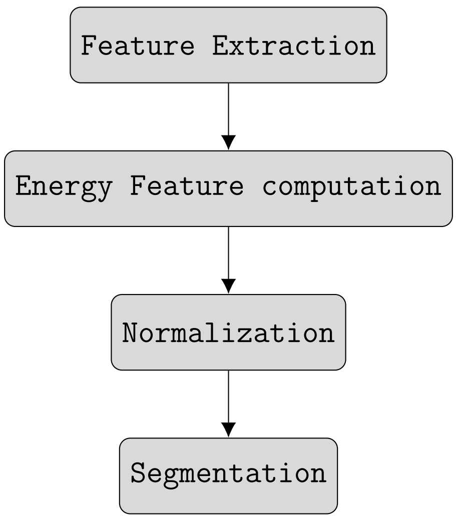
Understanding which feature dimension has the strongest discriminant power and which feature dimension has the weakest discriminant power can help to understand convolve the laws filters with image and the clustering process. E5-E5 has the strongest discriminant power because it is an edge-edge filter. The edge indicates the boundary between different object or texture. The frequency response in the edge is large which means a stronger discriminant power. The R5-R5 is a high pass filter which only passes pixels with high frequency. However, the number of type of pixels are relatively small. As a result, R5-R5 will have the weakest discriminant power.

1.2 B. Texture Segmentation

1.2.1 Abstract and Motivation

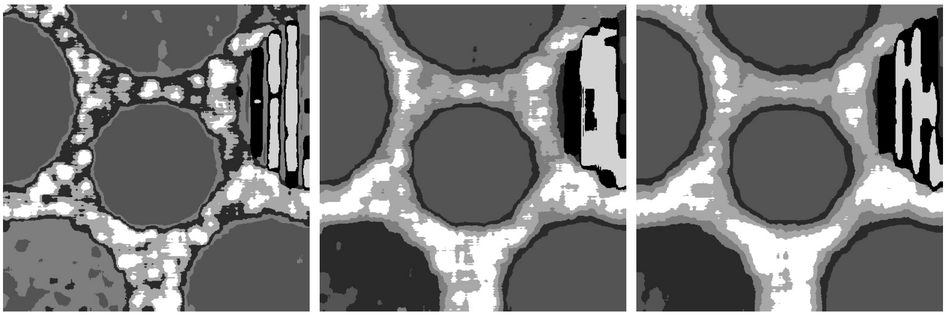
In human visual system, the texture segmentation is completed by seeing different colors and textures. A multi-channel filtering for visual information is applied in human visual system to segment different textures. People adopt the same idea to design a texture segmentation system for computers by using filters and clustering method. Texture segmentation is also one of important topic in computer vision and image processing.

1.2.2 Approach and Procedures



For feature extraction, apply the same Laws filter bank in the texture classification section to the input image to get twenty five gray-scale images. In the energy feature computation stage, use a window filter to compute the energy of each pixel of twenty five filtered images. A 25-D energy feature vector will be generated for each pixel. For energy feature normalization, energy feature of each pixel is normalized by the energy generated by L5-L5 filter. The reason is that L5-L5 filter does not have a zero-mean and it cannot extract useful features. Last, the k-means method is used to perform segmentation of textures.

1.2.3 Result



(a) Window Size 20 (b) Window Size 30 (c) Window Size 40

Figure 2: Texture Segmentation Out

1.2.4 Discussion

The figure 2 shows the output of texture segmentation. However, the performance is not that well. There are lots of missed label in the figure 2 because of high dimensional feature vectors.

1.3 C. Advanced Texture Segmentation Techniques

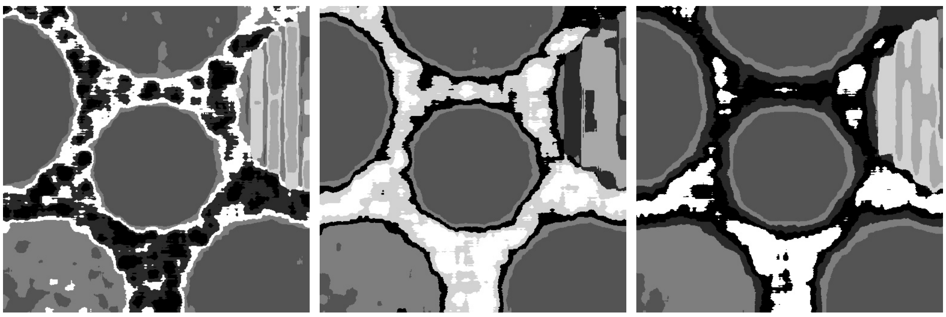
1.3.1 Abstract and Motivation

The performance of texture segmentation can be improved by applying PCA and holes filling method. This, section is focus on the PCA and hole filling method.

1.3.2 Approach and Procedures

As discussion in the section 1.1, the PCA method would reduce the computational time and improve accuracy by getting rid of the weak eign-pairs. The hole filling method will be able to fix the hole in the local region.

1.3.3 Results



(a) Window Size 20 (b) Window Size 30 (c) Window Size 40

Figure 3: Advanced Texture Segmentation Out

1.3.4 Discussion

Compared the figure 2 with the figure 3, there is a greate improvment by using the advanced texture segmentation techniques.

2 Problem 2: Image Feature Extractor

2.1 SIFT

2.1.1 Abstract and Motivation

Feature matching is an important topic in computer vision and image processing. The key-point detectors people developed before did not perform well when image is rotated and scaled. The Scale Invariant Feature Transform (SIFT) developed by Lowe successfully solved feature matching problem with changing scale and rotation.

2.1.2 Approach and Procedures

The SIFT algorithm used detector and descriptor to detect and describe keypoints [2]. For the detector, create a scale space of images which means to construct a set of progressively Gaussian blurred images and take differences to get a Difference of Gaussian (DoG) pyramid (similar to a Laplacian of Gaussian, 'LoG'). Then, find local extrema in the scale-space and choosen keypoints from the extrema. Use the descriptor to describe the important information of keypoints. For each keypoint, in a 16×16 window, find histograms of gradient directions and create a feature vector out of these. The following figure 4 shows the SIFT descriptor formation [2]:

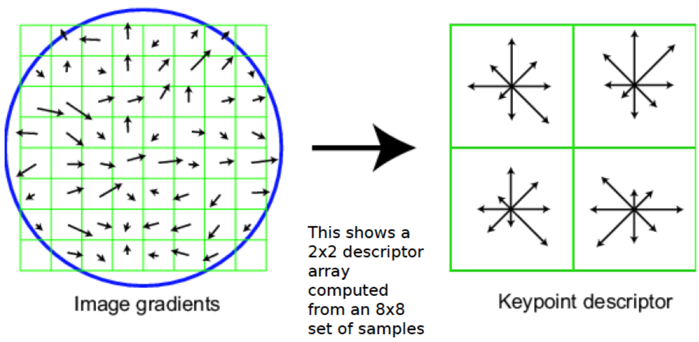


Figure 4: 3D Feature Vector Visualization

2.1.3 Result:

Read the paper "Object recognition from local scale-invariant features" by Lowe, David G in iccv Ieee 1999 and finish the question below.

- **From the paper abstract, the SIFT is robust to what geometric modifications?**
  The SIFT proposed by David is robust to translation, scaling, and rotation. In addition, it is partially robust to affine or 3D projection.

- **How does SIFT achieves its robustness to each of them?**
  Translation: Early researches about detecting features were robust in translation. They adopted corner detectors to extract key features. The SIFT algorithm constructs a set of progressively Gaussian blurred images and then takes differences to get a difference of Gaussian (DoG) pyramid. Determine the keypoints from finding local extrema in the DoG pyramid.
  Scaling: Changing in scale of images will cause feature detection fail in the previous researches. However, the SIFT algorithm achieves scaling invariant by building DoG pyramid and finding local minima and maxima in these pyramid to determine keypoints.
  Rotation: The invariance of rotation achieves by creating a scale space of images (construct a set of progressively Gaussian blurred images and take differences to get a DoG pyramid) and then finding local extrema in this scale-space to determine features.

- **How does SIFT enhances its robustness to illumination change?**
  To enhance the robustness of illumination change, the SIFT algorithm sets thresholding value at 0.1 maximum possible gradient value which is calculated by taking derivative of DoG pyramid to determine possible features.

- **What are the advantages that SIFT uses difference of Gaussians (DoG) instead of Laplacian of Gaussians (LoG)**?
  First, the Laplacian of Gaussian images are not scale invariant. The σ^2 in the demonimator is scale. We will get scale independence by removing σ^2. The difference of Gaussian adopts this idea to remove scaling effects.
  Second, the difference of Gaussian images are approximately same to the Laplacian of Gaussian. Also, DoG is less computationally intensive than LoG.

- **What is the SIFT's output vector size in its original paper?**
  There are 30,000 key vectors in the paper.

2.1.4 Discussion

The Speeded-Up Robust Features (SURT) developed by Bay in 2006 speed up the SIFT process by using block summing operations rather than convolutions. The SURF uses box filters to approximate DoG. Instead, the SIFT using DoG to approximate LoG.

2.2 B. Image Matching

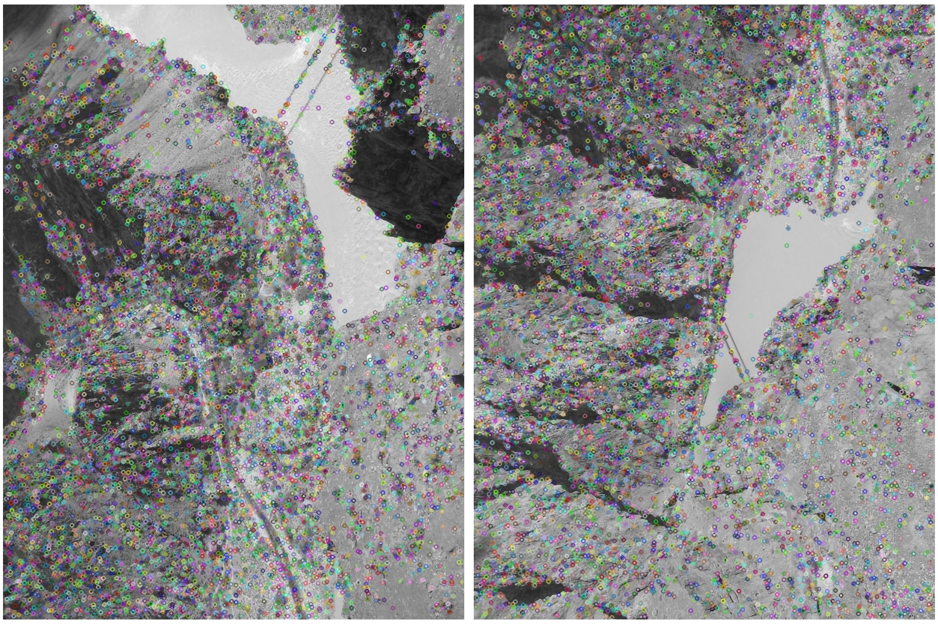
2.2.1 Abstract and Motivation

The major job for SIFT is to do the image matching. In this section, two river images are given to experiment with SIFT algorithm.

2.2.2 Approach and Procedures

Use the OpenCV library's SIFT tool to find the keypoints in the given image, and match these keypoints. The SIFT algorithm is discussed in the section 2.1.

2.2.3 Result



(a) River 1 Keypoints (b) River 2 Keypoints

Figure 5: Keypoints in the Given Image

The figure 5 shows the keypoints of river1 image and river2 image.

Figure 6: Keypoint with the largest scale pair

The figure 6 shows the keypoints pair in two images.

2.2.4 Discussion

The gradient magnitudes and orientations in a small window around can be calculated by following equations:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\Theta(x,y) = tan^{-1}(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y})$$

After this, generate a histogram of gradient orientation with 36 bins to describe each keypoint. In addition assign the dominant orientation which is the highest count in histogram as the orientation of the keypoint. In case of multiple peaks more than 0.8*peak, create a separate desciptor for each orientation. The two keypoints matched in the image have the close histogram of gradient orientation which means that the orientation of two keypoints is closed and also their value is closed too.

2.3 C. Bag of Words

2.3.1 Abstract and Motivation

In addition to application in image matching, the SIFT algorithm can be also applied in classifying images. The Bag of Words (BoW) proposed by Fei-Fei Li in 2005 is a image classification method in computer vision by treating image features as words.

2.3.2 Approach and Procedures

BoW contains two parts: feature representation and codebook generation. The SIFT algorithm is used in the feature representation to find locations of keypoints and generates descriptors of these keypoints. For codebook generation, perform k-mean clustering over all descriptors to generate codebook. Last, the testing image can be represented by the histogram of the codebook [3].

2.3.3 Result

The following chart shows the histogram of BoW for test image eight.

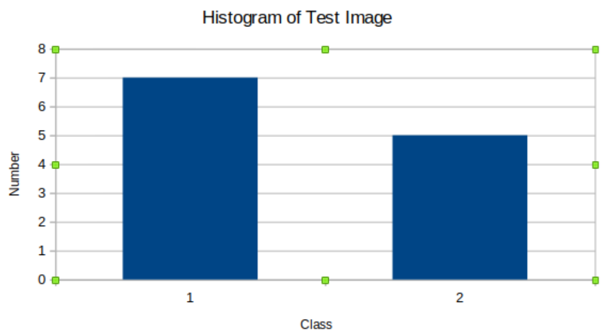


Figure 7: Histogram of the Bag of Words for Digit Eight Image

2.3.4 Discussion

From the figure 7 above, the digit eight is more likely to be classified as 0. Because, the training image 0's keypoints and their descriptors are more similar to 8's keypoints and descriptors. Their are only 2 classes and total 8 images available for training. The small dataset would cause missing classify of image. As a result, increasing the training data and include all digits in training will help to classify digits.

References

[1] C.-C. Jay Kuo *EE 569 Digital Image Processing: Homework 4* USC, Spring, 2019

[2] Lowe, David G *Object recognition from local scale-invariant features* iccv. Ieee, 1999

[3] Fei-Fei Li; Perona, P. *A Bayesian Hierarchical Model for Learning Natural Scene Categories* 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2. p. 524.