A Gabor Filter-Based Method For Handwritten Digits Recognition

Jing Liang[1] and Fengcan Zhu[1]
ECE Department, UCSD
9500 Gilman Drive La Jolla, CA 92093
jil884@eng.ucsd.edu f9zhu@eng.ucsd.edu

***Abstract*— In this report, we implemented the gabor filter to extract the features of some handwritten digits and produce a model from the extracted features. Then we apply the test data to test the produced model and examine the handwritten digits recognition rate**

I. INTRODUCTION

Handwritten digits recognition is very popular in recent years. Applications such that zip-code recognition, bank form processing are being widely used. It highly increase the productivity and bring convenient to our life. The whole process is scanning the digits to grayscale images and then converting to binary images. Then extract features from different classes and build a corresponding model according to training sets, then use test sets to do the decision making. In the past few years, several methods have been applied to do the digits recognition[1], like direct matching, zernike moments, zoning, geometric moment invariants.

Gabor filter has been used in many fields, such as target detection, character recognition, fingerprint recognition, face recognition, document analysis, image analysis and so on. An important point about Gabor filter is it possess the optimal localization in both spatial and frequency domain[2].Besides, Gabor functions are also closely related to the human visual system and texture interpretation, so it can be used to extract components according to different scales and orientations from 2D image.

II. GABOR FILTER

Gabor filter is a linear filter usually used for texture analysis and extract components corresponding to different scales and orientations from images. A 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal wave. It is defined by

$$g(x, y) = a(x, y)c(x, y) \tag{1}$$

where

$$a(x, y) = \frac{e^{-0.5(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2})}}{2\pi\sigma_x\sigma_y} \tag{2}$$

being the Gaussian component and

$$c(x, y) = cos(2\pi(F_x x + F_y y)) \tag{3}$$

being the sinusoidal component of the filter[1].

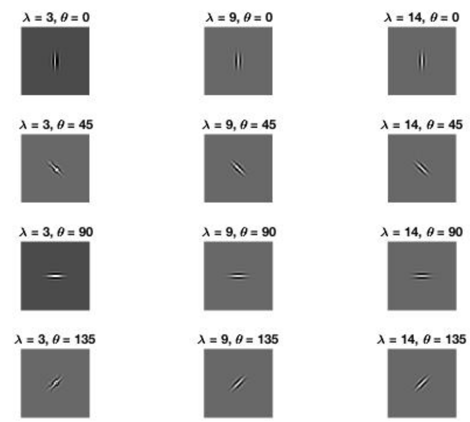

Fig. 1. Gabor filter bank in different wavelength and orientation

The variables, x and y are the spatial variables. The standard deviations (σ_x,σ_y) in (2) describe the size of the Gaussian envelope and define the scale of the filter along the spatial and spectral axes. (Fx,Fy) represents the frequency of the sinusoidal component and thus the center frequency of the filter in the 2D frequency domain. The orientation of the filter is defined as the unit vector from the origin to the center frequency (Fx,Fy) of the filter[1]. By changing the wavelength and orientation of the gabor filter, a Gabor filter bank is then composed to extract the features from the test dataset. The filter bank can be shown in figure1.We choose the orientation to be 0, $\pi/4$, $\pi/2$, $3\pi/4$, and the wavelength to be $\lambda = 3, 9, 14$(Since the range of λ is [2,$M/2$] while M is the size of image I). There are total 12 filters in the final filter bank.

III. FEATURE EXTRACTION

Gabor filters with different frequencies and orientations in different directions have been used to localize text regions. In this article, we will use Gabor filter to extract features in handwritten digits. Assume we create a Gabor filter banks with m different wavelengths and n different orientations. Now, we have total mn filters in the filter banks. We do the 2D convolution of image I with every filter in filter banks. Thus we have got mn outputs images G(i,j) with $1 <= i <= n$ and $1 <= j <= n$ which have the same size to image I. The phase information of G(i,j) can be taken as a feature, because it contains information about the edge locations and other details in image I. The amplitude of G(i,j) can be taken as a feature and it contains some oriented frequency spectrum in every local of the image I[2]. The square sum of the different frequency responses with the same orientation can also serve as a feature. The orientation in which the local has the maximum energy can be taken as a feature as well. We can apply one feature or several combinations of features. For every image I, if we use k features, the size of feature vector is $1 * (mnk)$. In this task, for every class in 0-9, we choose 200 samples, so the size of feature matrix for a single class is 200*(mnk). Totally, we have 10 matrix of the same size and we use classification method to build the model.

IV. EXPERIMENT

A. Dataset

The Dataset we use in this assignment is from **Mnist**. It contains about 6k training samples and 1k test samples of each class. Each image contains one handwritten digits from 0 to 9 and is chopped to size 28*28, so that the location of the digit is almost in the center. However, due to the computational limitation, we are going to use only 400 training samples and 200 test samples for each class (4000 training samples and 2000 test samples in total).

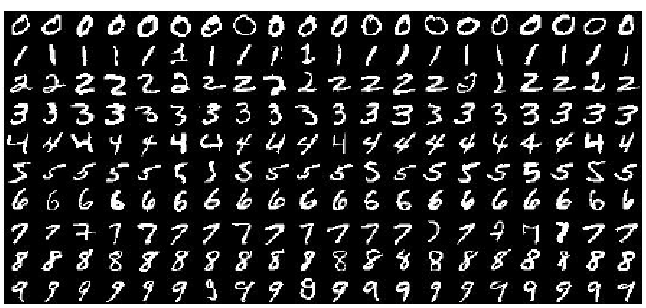

Fig. 2. Dataset Sample

B. Classification methods

We use KNN(K-nearest neighbors algorithm) to do the classification. In KNN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor. When k>1, the test result will be classified into the class which takes the most part. The algorithm is shown in the Figure3[3]. For example, the green circle represents the sample to be classified. It need to be classified into red triangle or blue square. If k equals 5, the green circle is classified to red triangle since the probability of classifying it into red triangle is 66% which is higher than that of classifying it into blue square, which is 33%.

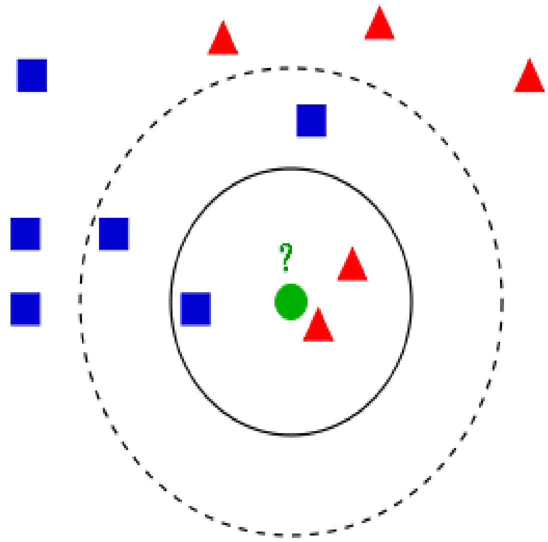

Fig. 3. K-nearest neighbors algorithm

C. Selection of parameters in Gabor filter bank

TABLE I

RECOGNITION RATE VS NUMBER OF ORIENTATIONS WITH K=5AND NUMBER OF DIFFERENT WAVELENGTH = 5

number of orientations	3	4	6	8
training set recognition rate	73.4%	75.9%	76.8%	77.9%
test set recognition rate	63.2%	66.5%	67.8%	69.2%

TABLE II

RECOGNITION RATE VS NUMBER OF WAVELENGTH WITH K=5 AND NUMBER OF ORIENTATIONS = 8

number of different wavelengths	3	4	5	6	7
training set recognition rate	74.3%	76.7%	77.9%	78.8%	80.1%
test set recognition rate	64.2%	67.1%	69.2%	70.0%	72.1%

The Gabor filter impulse we used has three parameters to tune: the number of orientations and the number of filters with different wavelength(Since the dynamic range of wavelength is quite narrow) and the value of k in k-NN algorithm. First, we keep the number of wavelength stable and increasing the number of orientations. The result is in TableI. Next, we keep the number of orientations stable and change the number of wavelength. The result is in TableII with number of orientations equals to 8. We can come up with the conclusion that the more the number of filter with different wavelength and orientations, the better the accuracy. Finally, we keep the number of orientations equals to 8 and number of different wavelength equal to 7 and explore how the value of k affect the results. From Figure4,we can see as k goes up, at first classification accuracy of training set and test set are increasing, then the accuracy increasing speed decreases. So we choose k=5.

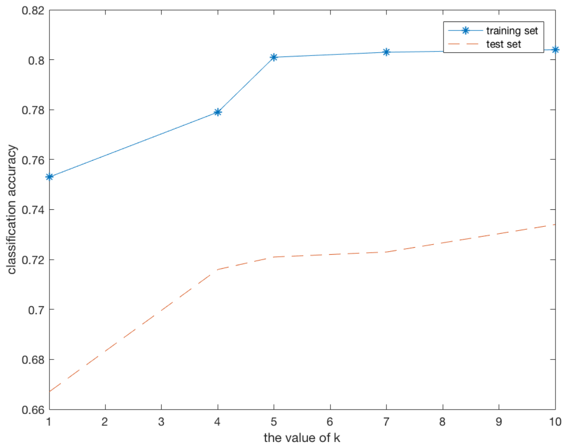

Fig. 4. the value of k vs classification accuracy

V. CONCLUSIONS

TABLE III

CONFUSION MATRIX OF FINAL RESULT ON TEST DATA

	0	1	2	3	4	5	6	7	8	9
0	83	0	1	0	1	2	8	2	2	1
1	0	91	1	1	1	1	1	2	0	2
2	0	0	74	4	1	4	5	3	2	7
3	0	0	6	73	0	12	2	0	6	1
4	0	14	0	0	49	1	5	4	2	25
5	0	0	6	21	0	66	4	1	1	1
6	8	0	3	1	2	0	65	1	5	15
7	2	3	3	4	8	1	9	55	1	14
8	0	2	3	2	1	4	3	0	79	6
9	1	2	0	2	3	1	3	4	7	77

TABLE IV

RECOGNITION RESULT WITH NUMBER OF WAVELENGTH = 7, K=5 , NUMBER OF ORIENTATIONS = 8 ON 1000 TEST DATA

digit	mostly misclassify as
1	7,9
0	6
8	9
9	8
2	9
3	5
5	3
6	9
7	9
4	9

Finally,by using the kNN classification and choosing 8 orientations and 7 different values of wavelengths and k =5, we achieve training accuracy 80.1% while test accuracy is only 72.1%.The confusion matrix of test set is shown in TableIII. Horizontal direction represents the actual class which the digit belongs to while vertical direction represents the predict class.From the TableIII we can see, digit '1' gets the best result with true positive rate equals to 91% while digit '4' gets the worst result with true positive rate equals to 49%. Mostly, the machine will confuse '4' with 9.The recognition result from best to worst are in TableIV.

However, it seems the recognition accuracy is not that high, we still have a lot to improve. Due to the computational limitation, our training sets is quite small compared with original given number of training samples thus we didn't make full use of the whole data. About extracting Gabor feature, we only use single feature instead using the combinations of features. In classification, we can use ANN(artificial neural network) if possible.

REFERENCES

[1] Trier, ivind Due, Anil K. Jain, and Torfinn Taxt. "Feature extraction methods for character recognition-a survey." Pattern recognition 29.4 (1996): 641-662.

[2] Zheng, Danian, Yannan Zhao, and Jiaxin Wang. "Features extraction using a Gabor filter family." Proceedings of the sixth Lasted International conference, Signal and Image processing, Hawaii. 2004.

[3] www.google.com

[4] P.Hu, Y.hao, Z.Yang, J.Wang. Recognition of Gray Character Using Gabor Filters. IEEE Comput. Soc. Part vol.2, 2000, pp.335-8 vol.2

[5] Y.Hamamoto, S.Uchimura, M.Watanabe, T.Yasuda, Y.Mitani, S.Tomita. A Gabor Filter-Based Methiod For Recognizing Handwritten Numerals. Pattern Recognition, Vol. 31, No. 4, pp. 395400, 1998

[6] X.Wang, X.Ding, C.Liu. Gabor filters-based feature extraction for character recognition. X. Wang et al. / Pattern Recognition 38 (2005) 369 379

[7] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces(Translation Journals style), IEEE Transl. J. Magn.Jpn., vol. 2, Aug. 1987, pp. 740741 [Dig. 9th Annu. Conf. Magnetics Japan, 1982, p. 301].