YTÜD 2001/4

ARAŞTIRMA MAKALESİ

THE USE OF STEERABLE FILTERS FOR HAND-WRITTEN CHARACTER RECOGNITION

Emir TUFAN AKMAN

İstanbul Üniversitesi,Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü Avcılar-İSTANBUL

Geliş Tarihi: 05.05.2001

EL-YAZISI KARAKTERLERİNİN TANINMASINDA YÖNLENDİRİLEBİLİR SÜZGEÇLERİN KULLANILMASI

SUMMARY

This paper proposes a new approach for hand-written character recognition, using steerable filters and neural networks where the former yields the local direction of dominant orientation and the latter, using this feature, recognizes the character. The input to the system consists of a 40x40 gray-scale distinct letter generated by 50 people for each hand-written digit. Based on a training set of 1000 and a set of 300 unknown hand-written digits 92 % correct recognition is provided.

ÖZET

Bu çalışmada yönlendirilebilir süzgeçler kullanılarak el-yazısı karakterlerinin tanınmasında yeni bir yaklaşım geliştirilmiştir. Yönlendirilebilir süzgeçler ile el-yazısı karakterlerine ait yerel baskın yönelim bilgisini içeren öznitelik vektörü elde edilmektedir. El-yazısı karakterleri bu vektörler ile karakterize edildikten sonra tanıma aşamasında sinir ağları kullanılmıştır. Test aşamasında, 40x40 boyutunda 1000 tane el-yazısı karakterine ait yerel baskın yönelim vektörü elde edilmiş ve tanıma işlemi için veri tabanı olarak kullanılmıştır. Daha sonra bilinmeyen 300 el-yazısı karakterine ait tanıma işlemi sonucu %92 doğru tanıma oranı elde edilmiştir. Böylece el-yazısı karakterlerinin tanınmasında yönlendirilebilir süzgeçlerin kullanılabileceği gösterilmiştir. Performans analizi yapılırken sinir ağı sadece bir araç olarak kullanılmıştır.

1. INTRODUCTION

Neural Networks have been used in character recognition in combination with different types of feature extraction methods. Several authors presented various approaches to extract features representing an image. Laine, Schuler and Girish [1] described a new method which includes an incremental strategy for recognizing characters based on the multiscale representation of wavelet decompositions. Castellano and Sandler [2] developed a new system for the same purpose using the Hough transform and a neural network. However, not much research has been carried into what may be called orientation techniques.

The orientation sub-band structures, known as steerable filters, have been used for image processing tasks, such as texture analysis, edge detection, image data compression, motion analysis, and image enhancement [3]-[5]. The reason for using the term ``steerable filter'' is that the orientation of the filter is steered to an arbitrary orientation according to the application requirement. These structures were first introduced by Freeman and Adelson [6] who studied the design and use of steerable filters. Also in [7], a novel alternative to current steerability approaches is proposed. This method is based on utilizing a set of polar separable filters with small support sample orientation information. Folsom and Pinter [8] proposed algorithms for quadrature disambiguation after an accurate method of edge and bar detection.

In this paper we develop a system for hand-written character recognition using a steerable filter and a neural network where the former yields the local direction of dominant orientation while the latter recognizes the character using this feature. The input to the system consists of a 40 x 40 image containing one hand-written digit and is processed using steerable filters. During this processing, the input is split into orientation sub-bands. After filtering, the outputs are multiplied by the so called interpolation functions which are expressed in terms of the steerable filter orientations. The energy measure for the pixels at the orthogonal axes in the image is defined as a squared sum of the sub-band outputs at different orientations.

The filters are steered to the local direction of dominant orientation hence maximizing the oriented energy. This process provides the characterization of the image by the local dominant orientation. Then this feature is fed into a neural network which in turn performs tie recognition task. Once the image has been pre-processed by steerable filters, then the testing is carried out using two different models of neural networks (i) the single layer perceptron with error correction (SLP-EC) and (ii) multilayer perceptron with back-propagation algorithm (MP-BP). The pre-processing enables a difficult global image recognition task to be fulfilled through a relatively easier local dominant orientation detection problem.

In this paper, we mostly concentrate on steerable filters as a pre-processing tool for the recognition problem. Although the recognition task is fulfilled through the use of neural networks, our intention in this study is to exploit them only as a tool, which is realized through the use of MATLAB Neural Network Toolbox.

2. STEERABLE FILTERS

Steerable filters basically provide directional edge detection since they behave as band-pass filters in a particular orientation [3], [9]. The edge located at different orientations in an image can be detected by splitting the image into orientation sub-bands obtained by basis filters having these orientations. In the structure shown in Fig. 1, $h^{\theta_i}(x,y), 1 \le i \le M$ is the rotated version of impulse response $h(x,y)$ at the filter orientation, θ_i, and $k_i(\theta_a)$, are called interpolation functions which control the filter orientations.

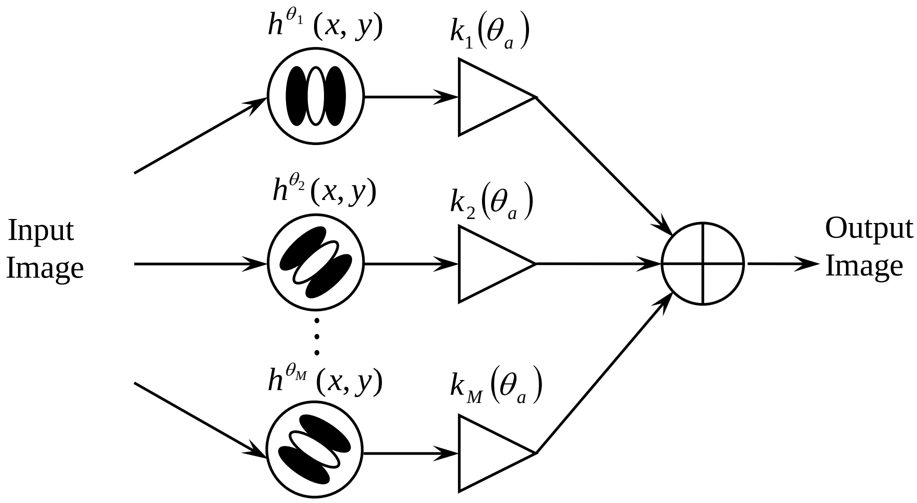

Figure 1. Steerable filter system block diagram.

Definition: $h(x,y)$ is said to be steerable if it can be expressed at an arbitrary rotation θ_a, as a linear sum of fixed rotated versions of itself, $h^{\theta_i}(x,y)$, i.e.

$$h^{\theta_a}(x,y) = \sum_{i=1}^{M} k_i(\theta_a) h^{\theta_i}(x,y) \qquad (1)$$

where $h^{\theta_a}(x,y)$ is the rotated version of $h(x,y)$ at θ_a direction and $k_i(\theta_a)$, are interpolation functions.

The following theorem ensures the steering condition [3]:
Theorem: The steering condition (1) holds for functions that can be expanded into Fourier series (i.e., periodic or compactly supported) in the following way:

$$h(\rho,\theta) = \sum_{n=-N}^{N} a_n^h(\rho) e^{jn\theta} \qquad (2)$$

where $\rho = \sqrt{(x^2 + y^2)}$ and $\theta = \arctan(y/x)$, if and only if there exists a set of interpolation functions $k_i(\theta_a)$ satisfying the matrix equation:

$$\begin{bmatrix} 1 \\ e^{j\theta_a} \\ \vdots \\ e^{jN\theta_a} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{j\theta_1} & e^{j\theta_2} & \cdots & e^{j\theta_M} \\ \vdots & \vdots & \ddots & [\vdots] \\ e^{jN\theta_1} & e^{jN\theta_2} & \cdots & e^{jN\theta_M} \end{bmatrix} \begin{bmatrix} k_1(\theta_a) \\ k_2(\theta_a) \\ \vdots \\ k_M(\theta_a) \end{bmatrix} \tag{3}$$

The number of nonzero coefficients $a_n^h(\rho)$ gives the minimum number of basis functions required for steering condition [3].

It is well known that the directional derivative operator is steerable [5], [6], [10]. We illustrate the use of steering constraint using the first derivative of the Gaussian function, $G(x,y) = e^{-(x^2+y^2)}$. The first x derivative is:

$$G_1^{0^o}(x,y) = h^{0^o}(x,y) = \frac{\partial G(x,y)}{\partial x} = -2xe^{-(x^2+y^2)} \tag{4}$$

which can be written in polar co-ordinates as:

$$G_1^{0^o}(\rho,\theta) = -2\rho e^{-\rho^2}\cos(\theta) = -\rho e^{-\rho^2}(e^{j\theta} + e^{-j\theta}) \tag{5}$$

It is obvious from (5) that the number of nonzero coefficients is two. The interpolation functions $k_i(\theta_a)$, can be found from (3) by:

$$e^{j\vartheta_a} = \begin{bmatrix} e^{j\theta_1} & e^{j\theta_2} \end{bmatrix} \begin{bmatrix} k_1(\theta_a) \\ k_2(\theta_a) \end{bmatrix} \tag{6}$$

The basis functions are spaced between 0^o and 180^o in order of symmetry. Therefore one basis function is oriented at 0^o and the other at 90^o. Substituting $\theta_1 = 0^o$ and $\theta_2 = 90^o$ into (6) we obtain

$$k_1(\theta_a) = \cos(\theta_a) \qquad k_2(\theta_a) = \sin(\theta_a) \tag{7}$$

Thus the first derivative of a Gaussian function at any rotation can be written as:

$$G_1^{\theta_a}(x,y) = \cos(\theta_a)G_1^{0^o}(x,y) + \sin(\theta_a)G_1^{90^o} \tag{8}$$

3. ORIENTED ENERGY

The oriented energy is used for analyzing the local orientation in an image [5], [11]. Knutsson and Granlund [12] devised a method for combining the outputs of quadrature pairs to extract a measure of orientation. Freeman and Adelson [3] also describe a method that makes optimal use of steerable quadrature filters for the same purpose. They measure the orientation strength along a particular direction θ_a by the squared output of a quadrature pair of band-pass filters steered to the angle θ_a.

The oriented energy is measured using *n*th derivative of Gaussian function, $G_n(x,y)$, and its Hilbert transform, $H_n(x,y)$, as

$$E_n^{\theta_a}(x,y) = \left[f(x,y) * G_n^{\theta_a}(x,y)\right]^2 + \left[f(x,y) * H_n^{\theta_a}(x,y)\right]^2 \quad (9)$$

where *f(x,y)* is the input image, $H_n^{\theta_a}(x,y)$ is the Hilbert transform of $G_n^{\theta_a}(x,y)$ and ($*$) represents the convolution operator. The expansion of (9) into Fourier series yields [3]:

$$E_n^{\theta_a} = C_1 + C_2\cos(2\theta_a) + C_3\sin(2\theta_a) + ...$$

Neglecting the higher frequency terms, the local dominant orientation $\theta_{a(\max)}$ and its strength *S* are obtained from $dE^{\theta_a}/d\theta_a = 0$ as:

$$\theta_{a(\max)} = \frac{1}{2}\arctan\left(\frac{C_3}{C_2}\right) \qquad S = \sqrt{\left(C_2^2 + C_3^2\right)} \quad (10)$$

The orientation map of an image can be constructed by calculating $\theta_{a(\max)}$ and *S* for each pixel in an image. Figs. 2 and 3 show the oriented energy plots with respect to angle at the point on the vertical and horizontal lines, respectively. This process can be used for each pixel in an image. As can be seen in these figures, the oriented energy reaches its maximum value at the orientation which is the orientation of the line. The oriented energy is obtained at 90^o and 0^o in Figs. 2 and 3, respectively.

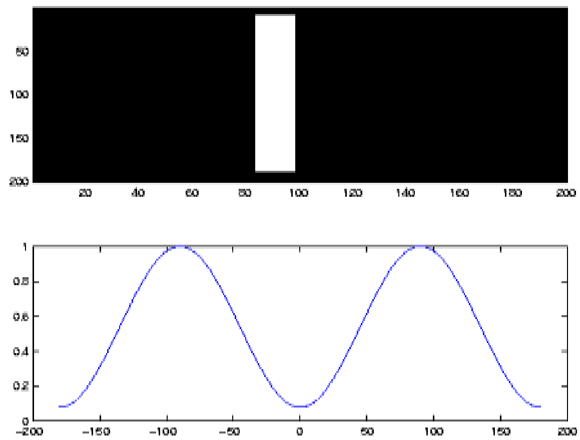

Figure 2. (a) Test image, **(b)** oriented energy with respect to angle for a pixel on the vertical line.

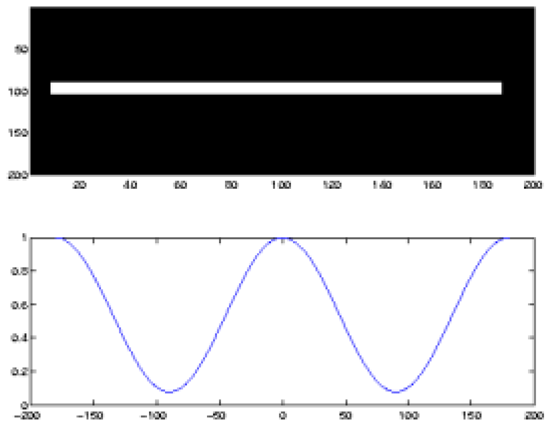

Figure 3. (a) Test image, **(b)** oriented energy with respect to angle for a pixel on the horizontal line.

As already noted in Section I, the pre-processing converts a difficult global image recognition problem into relatively easier local dominant orientation detection problem. We introduce orthogonal axes and detect dominant local orientation along these axes. These features are fed into a neural network for the recognition task. But the problem is that images have different sizes and they are placed in different locations. Laine, Schuler and Girish propose the method of accomplishing scale invariance [1]. In this method, each character is normalized in scale to fit a minimum bounding square. First the minimum bounding rectangle is constructed and the longest edge is identified. A minimum bounding square is allocated to match the length of the longest edge of the minimum bounding rectangle. Once embedded, a character may be shrunk or enlarged without distorting its original shape by the method of bilinear interpolation.

4. RECOGNITION WITH NEURAL NETWORK

In the recognition stage we will make use of a neural network. The testing will be carried out using (i) the single layer perceptron with error correction (SLP-EC) and (ii) multilayer perceptron with back-propagation (MP-BP).

Neural networks are composed of simple elements which are inspired by biological nervous systems. They have been trained to perform complex functions in various fields and find many applications such as pattern recognition, classification, identification, [2], [13]-[15].

The neurons of competitive networks learn to recognize groups of similar input vectors. Feature maps learn to recognize groups of similar input vectors in such a way that neurons which are physically close together in the neuron layer responding to similar input vectors.

A neuron can be described by writing the following pair of equations [16]:

$$u(q)=\sum_{r=1}^{R} w(q,r)\,p(r) \qquad\qquad \varphi\left[u(q)-B(q)\right] \tag{11}$$

where *p(1),...,p(r),...,p(R)* are the input signals while *w(q,1),...,w(q,r),...,w(q,R)* are the weights of neuron *q*. Here *B(q)* represents threshold function. An activation function,

$\varphi(.)$ provides for the limitation of the output of a neuron, *y(q)*. The main objective for image recognition is to find the best match of the input signals with the weights. *u(q),* $1 \le q \le Q$, are compared and the largest is selected for this purpose where *Q* is the number of neurons.

The accuracy of the algorithm depends on the number of epoch. Moreover, the success of the algorithm is critically dependent on how the main parameters of the algorithm, namely, the learning rate parameter η and activation function $\varphi(q)$. Unfortunately, there is no theoretical basis for the selection of these parameters. Nevertheless, the following observations provide a useful guide:

The learning rate η is a very sensitive parameter for the error correction as well as for the back-propagation algorithm. It does not only affect the rate of convergence of the learning, but also the convergence itself. If η is too small, the convergence of the network may be unacceptably slow. On the other hand, with too large learning rate, although the network may convergence faster to a stable solution, it may also diverge or oscillate and thus become unstable. The optimal value is not easy to reach since we need to adapt this parameter over the time of the training phase. Also the learning rate can be updated by considering recognition error as introduced in several studies [16]. In this study, the best results are obtained by updating the learning rate as $\eta = \eta/(1+N_E/100)$. Here N_E is the number of epoch. The starting value is chosen as 0.01 for single layer perceptron and 1 for multilayer perceptron.

The activation function defines the output neuron in terms of the neuron *q*. There are three types of activation function: (i) threshold function, (ii) piecewise linear function, and (iii) sigmoid function. We obtained the best result using the sigmoid function. The sigmoid function, whose graph is s-shaped, is by far the most common form of of activation function used in the construction of neural networks [16]. An example of the sigmoid function is the logistic function, defined by $\varphi(q) = 1/(1+e^{-\mu q})$ with μ is the slope parameter of the sigmoid function. By varying the parameter μ , sigmoid functions of different slopes are obtained. As the slope parameter approaches infinity, the sigmoid function becomes simply a threshold function. In our study, μ is chosen as 0.5.

4. CONCLUSIONS

Once the feature extraction task has been accomplished through the use of oriented energy concept, we need to use a tool to recognize the characters. In this study, two types of neural networks have been used: (i) the single layer perceptron with error correction (SLP-EC) and (ii) multilayer perceptron with back-propagation (MP-BP).

Different configurations have been used for both network structures. Our motivation in this choice has been that the system can be expected to vary with changing conditions and inputs while the system constantly learns. At this stage, we use the MATLAB Neural Network Toolbox.

For the MP-BP algorithm, we obtained the best results with one hidden layer having 10 neurons. Increasing the number of neurons in the hidden layer mostly resulted in lower recognition rate. The results for one hidden layer having 10 neurons with 4 and 10 outputs are shown in Table 1 for different number of inputs. With 120 inputs the best recognition rate obtained is 82% and with 60 inputs it is 92%. The performances of the methods introduced in [1] and [2] by using the data base which are used in this method are studied and with 60 inputs 90% and 85% correct recognition rates are obtained, respectively.

For the SLP-EC algorithm, the recognition rate is investigated according to the number of inputs and outputs. As can be seen from Table 2, a 85% recognition rate is achieved with 60% inputs for both and 10 outputs.

The work reported in this letter appears to be the only application where steerable filters are used for feature extraction in character recognition. The results obtained here are encouraging, however, there is a lot of room for improvement in all aspects of this study, i.e. the presentation of data to the steerable filter, the choice of steerable filter function, the implementation of the steerable filter, the presentation of the extracted orientation information to the neural network and NN itself. Further research is underway where these problems are being tackled.

Table 1. Multilayer perceptron with back-propagation where N_V , N_I, N_E are number of vectors, inputs and epoch, respectively.

Numbers of neurons	N_V	N_I	N_E	Correct Recognation %
120-10-10	6	120	3	30
			10	68
			80	85
120-10-4	6	120	9	75
			78	78
			204	82
60-10-4	6	120	10	72
			115	84
			195	88
			268	92

Table 2 Single-layer perceptron with error-correction.

Numbers of neurons	N_V	N_I	N_E	Correct Recognation %
120-4	6	120	7	62
			11	74
			110	77
60-4	6	60	3	43
			88	76
			418	85
120-10	6	120	118	76
			168	80
60-10	6	60	199	74
			311	85

KAYNAKLAR

[1] LAINE, A. F., SCHULER, S., GIRISH V., “Orthonormal wavelet representations for recognizing complex annotations”, Machine Vision and Applications, 6, 110-123, 1993.

[2] CASTELLANO, G., SANDLER, M. B., “Handwritten digits recognition using Hough transform and neural networks”, Proc. IEEE Int. Sym. on Circuits and Systems, Atlanta, USA, 3, May 1996, 313-316.

[3] FREEMAN, W. T., ADELSON, E. H., “The design and use of steerable filters”, IEEE Trans. on Patt. Anal. and Machine Intell., 13, 891-906, 1991.

[4] LAINE, A. F., CHANG, C. M., “De-noising via wavelet transforms using steerable filters”, Proc. IEEE Int. Sym. on Circuits and Systems, Chicago, USA, 3, May 1995, 1956-1959.

[5] KASS, M., WITKIN A., “Analyzing oriented patterns”, Comp. Vision Graphics Image Processing, 37, 362-385, 1987.

[6] FREEMAN, W. T., ADELSON, E. H., “Steerable filters” Topical Mtg. Image Understanding Machine Vision Opt. Soc. Amer. Tech. Digest Series, 14, 1989.

[7] YU, WC, DANIILIDIS, K., SOMMER, G., “Approximate orientation steerability based on angular Gaussians”, IEEE Trans. on Image Processing, 10(2),193-205, 2001.

[8] FOLSOM, T.C., PINTER, R.B., “Primitive features by steering, quadrature, and scale”, IEEE Trans. on Patt. Anal. and Machine Intell., 20(11), 1161-1173, 1998.

[9] TUFAN, E., “Multisubband structures and their application to image processing”, PhD thesis, South Bank University, 1996.

[10] HARALICK, R. M., “The digital step edge from zero crossings of second directional drivatives”, IEEE Patt. Anal. Machine Intell., 6, 58--68, 1984.

[11] MORRONE, M. C., OWENS, R. A., “Feature detection from local energy”, Patt. Recog. Lett., 6, 303--313, 1987.

[12] KNUTSSON, H., GRANLUND, G. H., “Texture analysis using two-dimensional quadrature filters”, IEEE Comput. Soc. Workshop Comp. Architecture Patt. Anal. Image Database Mgmt., 1983, 206-213.

[13] UTSCHICK, W., EIGENMANN, R., NOSSEK, J. A., “Hierarchical learning in multiresolution approach to neural networks”, 12th European Conf. on Circuit Theory and Design, Istanbul, Turkey, Aug. 1995, 327-330.

[14] KAHAN, S., PAVLIDIS, T., BAIRD, H., “On the recognition of printed characters of any font and size”, IEEE Trans. Patt. Anal. Mach. Intell., 9, 274-287, 1987.

[15] RAJAVELU, A., MUSAVI, M. T., SHIRVAIKAR, M. V., “A neural network approach to character recognition”, Neural Networks, 2, 387-393, 1989.

[16] HAYKIN, S., “Neural Networks”, Macmillan College Publishing Company, Inc., 1994.