Segmentation and Classification for Mixed Text/Image Documents Using Neural Network

Shinichi Imade Seiji Tatsuta Toshiaki Wada

Imaging System Dept. Olympus Optical Co.,Ltd.

Abstract

This paper presents a segmentation and classification method for separating a document image into printed character, handwritten character, photograph, and painted image regions. A document image is segmented into rectangular areas. Each of which contains a cluster of image elements. A layered feed-forward neural network is then used to classify each segmented area using the histograms of gradient vector directions and luminance levels. We obtained a high classification performance even with a small number of training samples. It is confirmed that the histograms of gradient vector directions and luminance levels are significantly effective features for the classification of the four kinds of image regions. Increasing the number of the discrimination areas improves the classification performance sufficiently even using a small number of training samples for the neural network.

1: Introduction

Techniques for dealing with various types of images are spectacularly progressing. High performance filing systems will be popular in a few years due to the development of high resolution imaging devices, high density storage apparatuses, and high speed microchips.

Data compression is one of the most important technologies for photographs and painted images because their data are more enormous than those of text images. The data compression technology has also been advanced. However, the efficiency of image compression depends on image types. In order to compress a document image more efficiently, therefore, we need to separate the document image into sub-images each of which consists of a single image type such as characters and photographs. Then we need to individually compress each sub image using an appropriate algorithm.

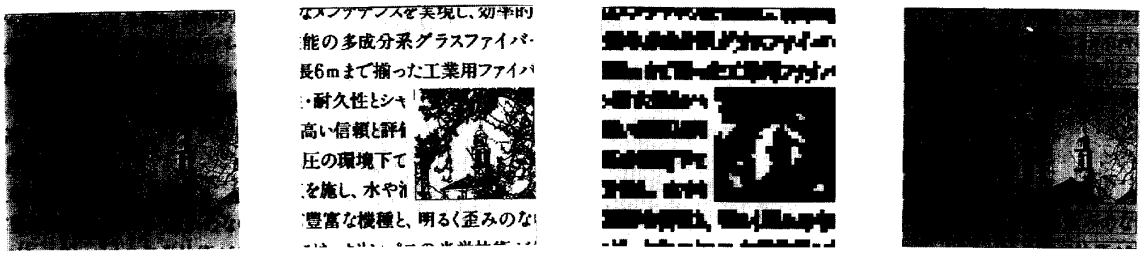
Many techniques for the segmentation and classification in mixed text/photograph documents have been proposed [1]-[4], but simple methods of classifying a document image into printed characters, handwritten characters, photographs, and painted images have not been reported. Painted image should be discriminated from photographs because the compression technique for photograph is not effective on painted images which contain higher frequency components. It is useful to classify characters into printed and handwritten for preprocessing of character recognition. Since traditional approaches to separate a document image into many types of image regions compare each separated region to many basic patterns, they require a large amount of computations.

This paper presents a new technique to segment and classify a document image into printed Kanji and Kana characters, handwritten Kanji and Kana characters, photographs, and painted images using a neural network. The proposed technique does not power and can realize precise segmentation and classification for mixed document images.

2: Segmentation

Fig1(a) shows an example of mixed document image which contains printed Kanji and Kana characters and a photograph. The aim of the classification is to separate the image into printed character areas and a photograph area. The first stage of our method have two steps: segmentation and classification. In the segmentation step, clusters of image elements of a document image are separated into rectangular areas. The segmentation process uses a binary monochromatic image which is obtained from the differential luminances of the original image by an appropriate thresholding [Fig.1(b)]. The binary image is then divided into a discrete blocks of 8×8 pixels, and every block image is reduced to one element. If a given 8×8 block contains at least one black pixel, this block is replaced by one black element. If the block does not contain any black pixels, it is replaced by white one [Fig.1(c)]. Isolated elements, which are considered to be noise, are removed by a spatial filtering. When at least 16 white elements continue in a vertical or a horizontal direction,

0-8186-4960-7/93 $3.00 © 1993 IEEE



(a) Input image (b) Converted binary image (c) Reduction (d) Result of Segmentation

Fig.1 Segmentation process

the elements are converted to a white line segment. Any other element than the white line segments is converted to black elements. Each region connected with black elements is enclosed by a rectangle. The classification is performed to each segmented area of the original image corresponding to a reduced rectangle shown Fig.1(d).

3: Feature Extraction

If all pixel values of the segmented areas are directly input to a neural network, it is difficult to classify the images because of a large scale network. In many applications, some significant features are first extracted which represent the relationship between images and their categories, and the images are then classified based on those features.

The classification technique proposed in this paper uses two combined features that are distributions of gradient vector directions and luminance levels. A block image to be classified is given as a square of N×N pixels in the segmented area. Let the luminance levels of a pixel at (i,j) be denoted by

$$y_{ij} \qquad (i, j: 0, 1, 2, \ldots, N-1). \qquad (1)$$

There is a pair of the gradients of a pixel in two orthogonal directions: horizontal and vertical. We define the bidirectional gradients Δh_{ij}, Δv_{ij}, by

$$\Delta h_{ij} = y_{(i+1)j} - y_{ij},$$
$$\Delta v_{ij} = y_{i(j+1)} - y_{ij}, \qquad (2)$$

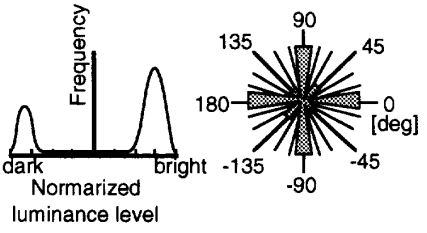
where i, j : 0, 1, 2, ... , N-2.

The gradient vector direction θ_{ij} for pixel (i,j) is given by

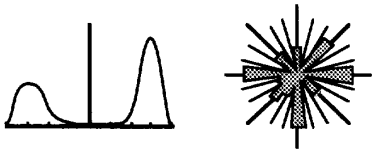
I $\quad \theta_{ij} = \tan^{-1}(\Delta v_{ij}/\Delta h_{ij}), \qquad (\Delta h_{ij} > 0)$

II $\quad \theta_{ij} = \tan^{-1}(\Delta v_{ij}/\Delta h_{ij}) + \pi \qquad (\Delta h_{ij} < 0).$

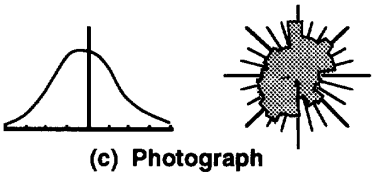
The gradient vector directions are calculated for all pixels in a block image, and the local sum for the histogram is computed at every quantized direction of 15 degrees. Therefore, the distribution is composed of 24 directional components. This distribution is normalized with the maximum value of the 24 components, and the resulting histogram of the gradient vector directions is given as one classification feature. Another feature is the histogram of luminance levels, where the histogram is constructed over each block image and is quantized by 32 components. These components are given by linearly dividing the range between the maximum and the minimum luminance levels and every local sum of these components is further normalized with their maximum value. The maximum luminance level is



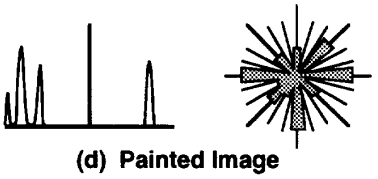
(a) Printed Character



(b) Handwritten Character



(c) Photograph



(d) Painted Image

Luminance level Gradient Vector Direction

Fig.2 Features of image classes

defined as the highest luminance level in histogram which excludes 5 % upper outliers, and the minimum luminance level is defined as the lowest luminance level in histogram which excludes 5 % lower outliers.

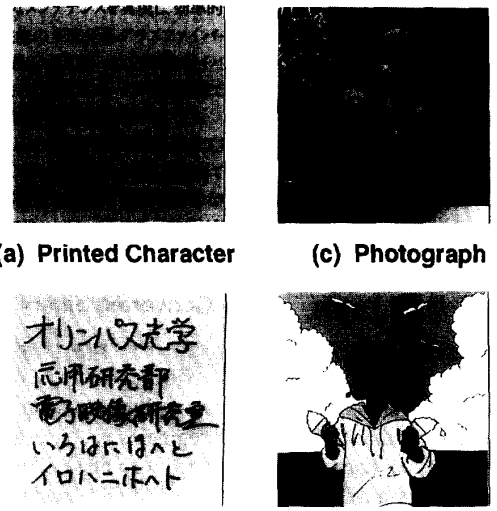
Fig.2 shows schematic drawings of the histogram of the gradient vector directions and the luminance levels. On the histogram of the gradient vector directions, printed characters have significant peaks in both horizontal and vertical directions. Painted image does not have significant peeks in both directions of the gradient vector. In handwritten character images, the histogram has same peaks but smaller than those of printed characters. The gradient vector directions of photograph cannot be characterized by any directions.

The histograms of the luminance levels of printed characters and handwritten characters have two peaks because of their binary gradation. Painted images have some peaks because of their discrete gradations. Photographs have relatively uniform distribution over the luminance range because of their continuous gradation.

4 : Classification

The segmented square areas are classified into the four kinds of image regions, printed characters, handwritten characters, photograph and painted image. In this experiment, small windows, named discrimination image blocks of 32×32 pixels, are selected randomly in the segmented area. The histograms of gradient vector directions and luminance levels are obtained in each discrimination image block.

These two histograms are input to a layered neural network (NN) which has already been trained with sample images. The NN consists of three layers: an input layer with 56 units whose number corresponds to the size of the feature vectors, a single hidden layer with 30 units, and an output layer with 5 units which are composed of the four image classes and the background of a document image. The training is performed by the error back propagation [5]. The training data, also called training vectors, are obtained from four sample images shown in Fig.3. The number of training vectors is 30 for each sample image, and the image blocks are randomly picked up in the samples. Each of training vectors consists of 24 components of the histogram of gradient vector directions and 32 components of the histogram of luminance levels that are calculated in an image block as described in the previous section. The desired output of the NN is given for every training vector. Only one element of the five output units is '1', others being '0'. The NN was trained with 150 training samples, namely 120 for the 4 image classes and 30 for the background of the image in Fig.3(a) and (b). The classification is performed for many discrimination blocks in a segmented area and its



(a) Printed Character (c) Photograph

(b) Handwritten Character (d) Painted Image

Fig.3 Image for training

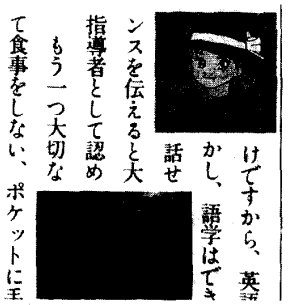


Fig.4 Image for separation

image class is determined by the majority of the results of the discrimination by the NN.

5 : Simulation

Computer simulations to segment and classify the mixed document image as shown in Fig.4 were performed using the proposed method. Fig.5(a) depicts the result. In this experiment, we used different training data samples from those shown in Fig.4. The image shown in Fig.4 does not contain handwritten characters, although the test image for the training data include handwritten characters. Fig.5(b) shows another result which was obtained using the image for the training data. Figs.5(a) and 5(b) indicate the relationship between the number of the discrimination image blocks in a segmented area and the classification performance which corresponds to the rate of correct recognition. 1000 experiments of the classification were carried out on each class of images. The locations of the discrimination blocks in each segmented area were selected

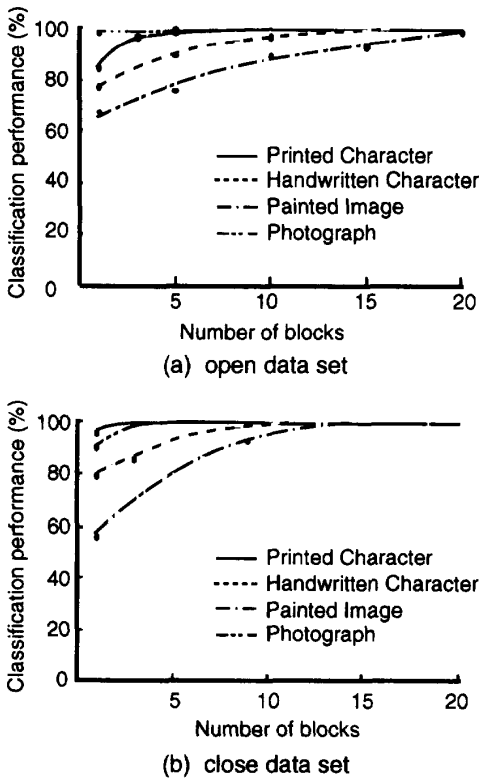


Fig.5 The classification performance of image classes

so as not to make piles.

From Figs.5(a) and 5(b) show that, as the discrimination blocks increase, the algorithm performs a better classification for all image types and leads to asymptotically 100 % correct classification. The classification performances for printed characters and photographs reach nearly 100 % even with only 5 or 6 blocks. More discrimination blocks for handwritten characters and painted images are required to achieve accurate classification. The classification performances for the open data with respect to the number of the discrimination blocks show almost same results as the close data, and they reach nearly 100 % in only 30 training data for every class.

A satisfactory classification for all kinds of the images can be realized by increasing the discrimination blocks and accepting the majority decision. It is shown that the extracted features in this method are satisfactory for the image classes because it can attain high performances of classification by using a small number of training data samples and setting not to be large numbers for the discrimination blocks.

The advantage of the combined feature, gradient vector directions and luminance levels, is clarified by comparing the simulations in which the classification process is executed with only one feature and their performance are examined. The discrimination blocks are selected in a segmented area and the tests are iterated in 1000 times. The same image as Fig.5(a) is used as the target to classify.

In the classification of the segmented areas only with the histogram of gradient vector directions of a single discrimination block, we find that the classification performance for the image of handwritten characters is inferior to the other image types and is below 50 % as shown in Fig.6(a).

Only with the histogram of only luminance levels, the classification performance for printed characters and handwritten characters are inferior to others and is below 50 %. With the combined features of the histograms of gradient vector directions and luminance levels, the correct classification performed more than 50 % for any kind of images. From the principle of majority decision, if the

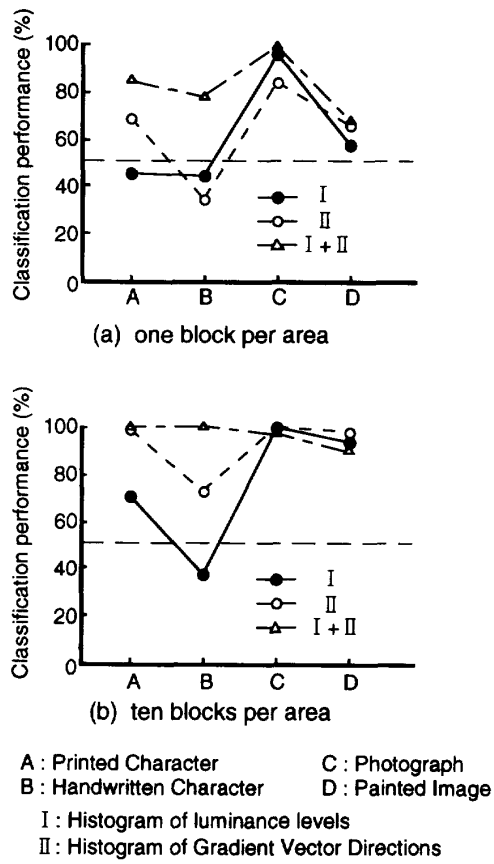


Fig.6 The classification performance for kinds of features

classification performance for a single discrimination block is over 50 %, the 100 % classification performance can always be achieved by increasing the discrimination blocks. But the classification processed with one of the two features cannot be used in practice. Because there exist some classes of images for which the increase of the discrimination blocks cannot lead to 100 % classification performance. Fig.6(b) shows the classification performances which are examined using one of the features with 10 discrimination blocks. In the classification for the handwritten character images using the histogram of luminance levels, the performance with 10 discrimination blocks is lower than the previous one with a single discrimination block.

Table 1 indicates the reason that the ratio of the correct classification for the handwritten images is lower than one of the incorrect classification. Increasing the number of the discrimination blocks, the classification performance of the handwritten images is getting worse. On the other hand, there are two cases that the classification performance is below 50 % with a single discrimination block but is improved by the majority decision of 10 discrimination blocks, one for printed character with the histogram of luminance levels and the other for handwritten characters with the histogram of gradient vector directions. This is because the ratios of the correct classification for both cases are the highest even with the only one feature.

Number of blocks	Classification result (%)				
	A	B	C	D	E
1	48.0	45.5	0	6.5	0
10	63.0	37.0	0	0	0

A : Printed Character
B : Handwritten Character
C : Photograph
D : Painted Imade
E : Background

Table 1 The classification result of Handwritten Character for the number of blocks per area using histgram of luminance levels

We examined the effect of discrimination block size to the classification performance. Table2 shows the classification performance for each kind of images with a single discrimination block of different sizes: 32×32, 16×16, and 8×8 pixels. In this configuration the combined feature was used and the classification was executed under the same condition of the previous experiment. The 32×32 block is nearly the same size of one character. The performance of printed character and handwritten character are worst in 8×8 block size but the effect of discrimination block size is not large. Since the 8×8 block size is much smaller than ordinary characters of documents, the areas containing a large character, such as title area, can classifiedwithout changing the block size, which is a suitable for an area of ordinary characters. We can use the same combined features in most cases once we determine a suitable size of discrimination blocks.

Block size (pixels)	32×32	16×16	8×8
Printed Character	84.5	75.0	45.4
Handwritten Character	77.7	64.4	40.0
Photograph	98.8	95.0	84.4
Painted Image	67.7	63.9	60.6

Table 2 The classification result (%) of each kind of image with different block size using the combined feature

6 : Summary

This paper described a new approach to segmentation and classification for the mixed document images consisting of printed characters, handwritten characters, photographs, and painted images. The two features, the histograms of gradient vector directions and luminance levels, are first extracted, and the classification is then carried out with a layered feed-forward neural network using these twofeatures simultaneously. We obtain the classification performance of nearly 100 % for the five image classes through the majority decision of the results given by the neural network which is trained using only 30 sample data samples for each class. This method can be applied to documents with several sizes of characters without changing discrimination block sizes and the two classification features.

References

[1] J.Weszk, et al., "Comparative study of texture measures for terrain classification," IEEE Trans. Syst. Man & Cybern., SMC-6,4,pp.269-285 (1976).

[2] Haralick,R.M. , "Edge and Region Analysis for Digital Image Data," Computer Graphics and Image Processing, Vol.12, pp.60-73 (1980).

[3] S.Tsujimoto and H.Asada, "Understanding multi-articled documents," Proc. 10th Int.Conf. Pattern Recognition(Atlantic City, NJ), pp.551-556 (1990).

[4] F.M.Wahl, K.Y.Wong, and R.G.Casey, "Block segmentation and text extraction in mixed text/image documents ,"Computer Graphics and Image Processing ,Vol.20, pp.375-390 (1982).

[5] D.E.Rumelhart, G.E.Hinton, R.S.Williams and the PDP Research Group, "Learning Internal Representation by Error Propagation", Parallel Distributed Processing, Vol.1, pp.318-362, MIT Press (1986).