

Image Pattern Recognition Based on Examples— A Combined Statistical and Structural-Syntactic Approach

Jiang Gao¹, Xiaoqing Ding¹, and Jing Zheng²

¹Institute of Image and Graphics, Department of Electronic Engineering,
Tsinghua University, Beijing, 100084, PRC
gj@ocrserv.ee.tsinghua.edu.cn

²Speech Technology and Research Lab, SRI International,
333 Ravenswood Avenue, Menlo Park, CA 94025

Abstract. An application of combined statistical and structural-syntactic approach in Chinese character recognition is presented. The algorithm adopts a structural representation for Chinese characters, but in the classification and training process, the structural matching and parameter adjustment is conducted in a statistical way. Different from the conventional *structural* approaches, in this system, only a few predefined “knowledge” is required. In most cases, knowledge acquisition is simplified to “memorization” of *examples*, and the parameters for classification can be refined using statistical training. In this way it avoids the main difficulties inherent in the implementation of classification systems based on structural features. Compared with conventional *statistical* algorithms, the algorithm is based on a structural model of image patterns, so it has approximately all the advantages of structural pattern recognition algorithms. A prototype system has been realized based on this strategy, and the effectiveness of the method is verified by the experimental results.

1 Introduction

As the situation in general area of pattern recognition, in Chinese character recognition^[1-2,4], there are two main trends: the *statistical approach* and the *structural-syntactic approach*. Though the combined statistical and structural approach has been widely used in online handwritten character recognition and offline Latin handwriting recognition, in the area of offline Chinese character recognition, due to the complexity of the problem, it is very difficult to apply the similar ideas. In this paper, a combined statistical and structural approach is proposed for offline Chinese character recognition.

Since there are many *advantages* associated with the statistical approach, such as a simple and objective algorithm structure, based on firmly established elements of statistical decision theory, easy to train, etc., the statistical approach has been used widely in practical systems.

Unfortunately, the statistical approach does have some *disadvantages*. For example, they are vulnerable to:

- Complex backgrounds, such as in the case of video information retrieval. Statistical algorithms usually require pre-segmentation of object-of-interest from backgrounds;

- Deformable pattern recognition tasks. For highly deformable patterns, it is difficult to train a statistical classifier to recognize all the deformed versions of image patterns with high fidelity.

On the other hand, the structural-syntactic approach has *advantages*^[1,3,4] in the above aspects, but its *disadvantages* are usually overwhelming. Some of them are:

- Designing a successful structural-syntactic “mechanism” is difficult—only human beings have this mechanism, and even human beings do not know how to realize this mechanism in a computer;
- Existing implementations are very complex, and hard to learn—just as above, only human beings know how to learn.

If we can combine these two approaches, the advantages of both approaches can be utilized to enhance the accuracy and robustness of the classification systems, while reduce the difficulties in implementations.

In this paper, a novel algorithm for recognition of Chinese characters is proposed using combined statistical and structural-syntactic approaches, and a prototype system, which can classify over 7,000 Chinese character, is realized based on this algorithm. The new algorithm has the following features:

1. A simple, objective algorithm structure, which is typical of the statistical-based approach;
2. Learning is accomplished by statistical training;
3. Structural representation of the image patterns;
4. Except for a little pre-defined *knowledge*, knowledge acquisition is simplified to “memorization” of *examples*.

We will describe the whole strategy and algorithm in section 2 to section 4; section 5 presents some experimental results using our prototype system; and finally, in section 6, some conclusions are given.

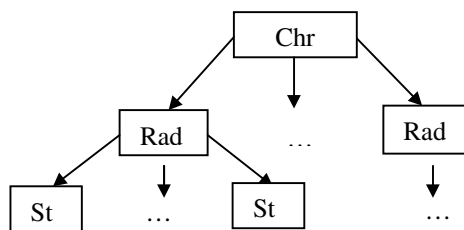
2 Knowledge Representation

One essential feature of the algorithm proposed here is that it is a hybrid of structural representation of image patterns and a statistical method for classification and training. The structural representation of image patterns is important since it endows many advantages, as described above, to the classification system.

2.1 Internal Structural Representation of Chinese Characters

A well-known structural representation of Chinese characters decomposes characters into radicals, and radicals into strokes^[1,4], as shown in Fig.1.

The reason for adopting this hierarchical representation of each character is that it is concise and is similar with the conceptual representation of Chinese characters by human beings.



Chr: Character Rad: Radical St: Stroke

Fig. 1. The structural representation of Chinese characters.

2.2 Knowledge Acquisition by a “Memorization” Process

In the conventional structural-syntactic approaches, knowledge acquisition is a very complex and difficult task. If we can replace knowledge by the memory of some quantities, then the implementation of the classification system will become much easier.

In this paper, we use the *online* method^[5,6] to input Chinese characters—we call these characters as “*examples*”, and several hundreds of radicals, to the classification system. After the input process, each “*example*” is recognized and decomposed into radicals, and each radical is decomposed into strokes automatically, by using the online character recognition algorithm developed by our Lab. The precision of this automatic classification and decomposition process can be very high if some constraints for character input are used. Then, the position and size of each decomposed radicals and strokes are stored in a database. We call this process the “*memorization*” process. In our system, only one *example* of each character is necessary.

Of course, only this memorized *examples* is not enough for classifying handwritten Chinese characters. A few “*knowledge*” is still required. We call these “*unreducible*” knowledge the “*primitive*” knowledge.

3 The “Primitive” Knowledge

As pointed out in the last section, pure memorized quantities may not be enough for the complex perception tasks. In major factors of human perception processes, there are surely something that should be called “*skills*”. We use the “*primitive*” knowledge to represent these skills.

In this paper, the “*primitive*” knowledge include a definition of standard strokes, which are in the lowest hierarchy in the structural representation of Chinese characters, and the ability of the system to locate and recognize these standard strokes

from complex backgrounds. Though this is a difficult task itself, but compared with the direct recognition of thousands of Chinese characters, each with different, complex

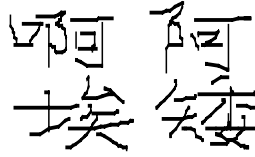


Fig. 2. Examples of the online-input characters.

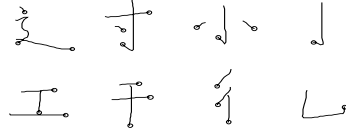


Fig. 3. Examples of the online-input radicals. They have been decomposed into strokes, as marked by “o”.

shapes, this *unreducible knowledge* requirement is much easier to meet. Our source codes for this task is less than 1,000 lines.

3.1 Stroke Descriptions

First, we define 17 standard strokes as the basic components of Chinese characters and radicals. These standard strokes are listed in Table 1.

Table 1. The standard strokes.

For each standard stroke, a context-dependent model for the stroke shape, namely *shape generation model* (SGM) is proposed. In SGM, we represent each stroke by a directed state transition diagram, as shown in Fig.4.

By modeling the strokes using the directed state transition diagrams, we assume that each stroke is composed of several line segments in the character image. We call these line segments *image cells*. We will describe how to segment a character image into image cells briefly in section 5. The image cells in a particular stroke can be connected or disconnected. The relation “connected” is represented by a transition between the two states that correspond to image cells. The “staying on” transitions, which are represented by self-loops on each state, are incorporated to compensate uncertainties in the segmentation of basic image cells, especially for arcs with great shape distortions which are common in handwritten characters.

The main attributes of the states in the state transition diagram are the directional properties of the image cells that are components of a particular stroke. The transition

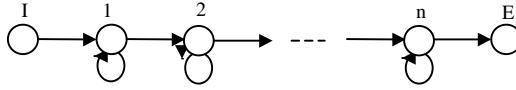


Fig. 4. The state transition diagram representation of shape generation models. I is the initial state, E represents the ending state. 1,2,...,n are intermediate states.

rules, which guide the transitions between states, describe the constraints in a particular shape, and also include possible distortions and disturbances, such as broken strokes.

The advantage of this shape generation model, as compared with the conventional *static* shape matching methods for stroke extraction, is that it deals with arcs, which are common in handwritten character shapes, and various distortions, including those caused by the skeleton extraction operation more effectively.

In Fig. 4, state “I” represents the initial state, which are initial image cells corresponding to the detected *attention focus areas* (AFA’s), which are described in the next section. Since initial image cells need not necessarily correspond to the starting or terminating line segments of strokes, in the shape generation process they are extended in two directions under the guidance of transition rules. In the following sections, we also use *initial cells* to refer to initial image cells or AFA’s.

3.2 A Selective Attention Approach for Stroke Extraction

Based on the SGM’s of strokes defined as above, we propose the following *selective attention* approach for stroke extraction from complex backgrounds.

The algorithm is composed of two steps: First, attention focus areas, i.e. AFA’s are detected; second, based on SGM’s of strokes, a *dynamic growing* procedure is applied to group image cells into candidate 2D stroke shapes.

Various features can be used to detect AFA’s. In experiments, we have simply used positional and directional relations between image cells, as well as the directional features of individual image cells to detect AFA’s. Other features can also be used. This flexibility of applying multiple features to detect multiple AFA’s guarantees the reliability and robustness of the shape extraction result, and is one advantage of the selective attention approach proposed here.

It is well known that the accuracy of stroke extraction is limited due to noises and uncertainties, no matter how much knowledge are used. In fact, we do not expect our algorithm to extract correct strokes all the time, we only require it to extract *possible* strokes, based on simple shape criterions. It is essential for the algorithm to extract *all* the possible strokes, as shown in the experiments, so we have relaxed some criterions to take into account some confusing conditions.

4 Statistical Approach for Structural Matching

Except applying the algorithm of section 3 to extract candidate strokes, other process of character recognition should be conducted in a straightforward and objective way, based on the memorized *examples*.

An optimal structure-matching algorithm for Chinese characters based on various statistical quantities describing the structural features is proposed in the following. In the structural matching procedure, the stroke components of the memorized *examples* are compared with their possible counterparts extracted from the character images. For this purpose, we need a statistical model to make comparisons and decisions.

4.1 The Statistical Model

In our algorithm, various statistical measurements of structural features are obtained and described using probability models. An example is given in Fig. 5. We measure the size of individual strokes using the quantities H, W , and relative position of stroke1 and stroke2 using LH and LW , respectively. Each of these quantities are Gaussian mixtures-modeled, and the parameters of these models, which are initially determined empirically based on the “*examples*”, can be refined by training, as mentioned in the following sections.

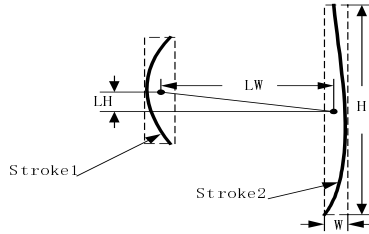


Fig. 5. An example of statistical measurements of sizes and relative positions of strokes.

4.2 Structural Matching

The matching between extracted strokes and those of the prototype *examples* is typically a many-to-one matching problem, so some effective *search* methods have to be applied to determine the optimal structural matching between strokes.

Compromising between accuracy and speed, we apply the *beam-search* method to select the optimal combinations. The procedure is: 1. Extract the candidate strokes from the input image. 2. Compute the statistical quantities of extracted strokes. 3. Compare these quantities with the corresponding data of the prototype models, i.e., the measurements obtained from the “*examples*”.

In this process, the strokes are added one by one, and at each step, only the top N combination of strokes with the highest matching scores are reserved until the next step. N is selected empirically. In this way we are able to control the computational complexity, while still maintain accuracy in most cases.

The procedure is conducted from one radical to another. We regard the extracted radicals as “macro-strokes”, i.e., compute the statistical quantities of these radicals the same way as we do for the ordinary strokes. Based on this “extended” stroke model, the structure matching process proceeds according to the procedure described above.

4.3 Learning: Training

One advantage of adopting the statistical approaches for structural matching is that, the *learning* process can be easily realized using statistical *training*, a conventional procedure in statistical pattern classification.

In our system, the data of the pre-memorized “examples” are used as initial parameters, and statistical measurements of the size/position of structural features from the input images are accumulated. These data are used to adjust the parameters for structural matching, such as the parameters of Gaussian mixtures. This supervised training is a typical EM process.

5 Simulations

We are interested to test the simple structure of the proposed strategy in some typical cases that are difficult to treat by either the statistical approach or the structural approach along. In this paper, we apply the algorithm to the recognition of a class of deformable image patterns, the handwritten Chinese characters. The procedure is given in Fig. 6.

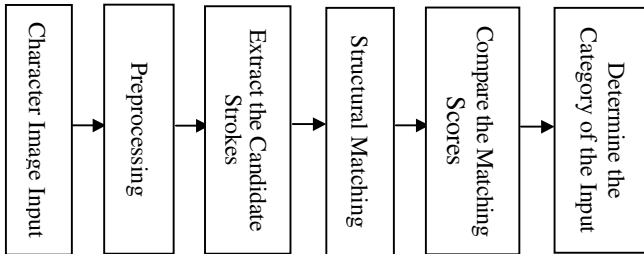


Fig. 6. The procedure of handwritten Chinese character recognition.

The different parts in the procedure are described below.

5.1 Preprocessing

The preprocessing procedure includes noise filtering using morphological filters, skeleton extraction, and most importantly, segmentation of the image skeleton into *image cells*, which are referred to in section 3. The image cells are defined as the line

segments with uniform directional attributes. To extract the image cells, we detect the intersections on the skeleton, and compute the curvature at each point on the skeleton. It turns out that the segmentation results of our algorithm are satisfactory, being good approximations to the results of human judgements.

5.2 Recognition of Deformable Characters

There are 7920 “*examples*” of handwritten Chinese characters in our database. For each input character images, we compare, or match the “structure” of this character image with each of the *examples* in the database, and the results of this structural matching are represented by the matching *scores* ($\in [0,1]$). The category of the input character are determined by comparing their “scores”.

When comparing the “structure” of input character image with an *example* in the database, the structural matching proceeds from one radical to another, using the algorithm described in section 4.2. An example of this procedure is given in Fig.7 to Fig.9. The candidate strokes in a particular radical are extracted from the character image using the algorithm described in section 3.2. The score of the “best match” is the final matching score of the character example with the input image.

Fig.7 to Fig.9 give the intermediate structural matching results for character 阿 with an input character 阿. The characters that give the highest matching scores for the input character 阿 are given in Table 2.

Table 2. Characters with the highest matching scores for an input image of character 阿.

Characters	阿	阳	阻	陌	河
Scores	0.7774	0.4639	0.3790	0.3311	0.3259

6 Conclusions

An algorithm for image pattern classification is proposed. The main feature of this classification algorithm is that it tries to combine the advantages of both the statistical and structural-syntactic approaches, while minimize the difficult parts in both of these two conventional approaches for pattern classification.

Compared with the conventional statistical approaches, the strategy proposed here makes use of the structural features of images, so is expected to give better results and flexibility in some real applications, esp. the extraction of deformable image patterns from complex backgrounds. In some way, the combined statistical and structural approach for image pattern classification can be viewed as a natural extension of the methodology for 1-dimensional signal classification, such as speech recognition. In our method, we have chosen the strokes as the primitive structural elements, or basic cells, since strokes are relatively simple, and correspond to the basic structural components in Chinese characters, just as the role of phonemes in speech models. The position/size relation model for a particular character in our system corresponds to the model for words, such as HMM, in speech recognition, and both of these models can be trained in a statistical way.

Future works on this direction include:

1. More accurate modeling of strokes, so the strokes can be extracted more accurately; and more suitable selection of statistical models for structural matching.
2. Accelerating the present algorithm by incorporating more constraints in the procedure of structural matching.
3. Conduct experiments on extraction of characters from complex backgrounds, such as in the application of video/image information retrieving, using this algorithm; and apply this strategy to the classification of other deformable image patterns in complex backgrounds.

References

1. Hildebrandt, T.H. and W. Liu, Optical recognition of handwritten Chinese characters: advances since 1980. *Pattern Recognition*, Vol.26, No.2, pp. 205-225, 1993.
2. F. Kimura, et al., Improvement of Handwritten Japanese Character Recognition Using Weighted Direction Code Histogram, *Pattern Recognition*, Vol.30, No.8, pp. 1329-1337, 1997.
3. Rocha, J. and T. Pavlidis, A shape analysis model with applications to a character recognition system. *IEEE Trans. PAMI*, Vol.16, pp. 393-404, 1994.
4. Hong Dong, Xiaoqing Ding and Youshou Wu, The ARG description of Chinese characters and a radical extracting method based on this description. *The Chinese Information Journal*, Vol.1, No.3, pp. 47-52, 1987.
5. C.C.Tappert, The state of the art in on-line handwriting recognition, *IEEE Trans. PAMI*, Vol.12, No.8, pp. 787-808, 1990.
6. F.Nouboud and R.Plamondon, On-line recognition of hand-printed characters: survey and beta tests, *Pattern Recognition*, Vol.23, No.9, pp. 1031-1044, 1990.

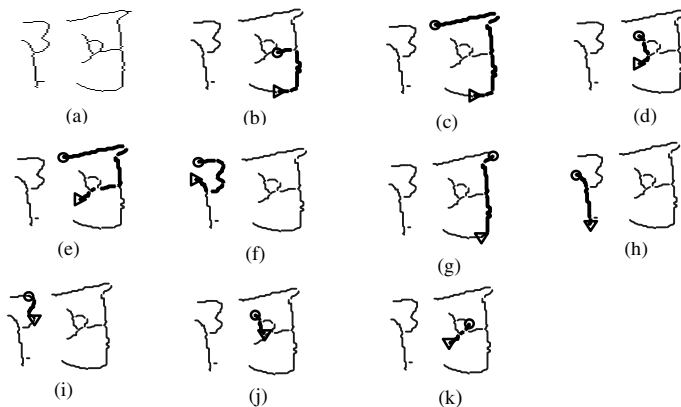


Fig.7. (a) The skeleton of input character image. (b)-(f) Some extracted candidates for stroke ㇇. —: the skeleton of the extracted strokes. - - : the skeleton of the input character. (g)-(k) Some extracted candidates for stroke ㇏.

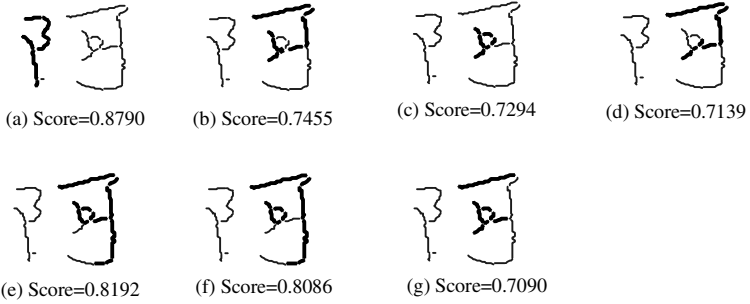


Fig.8. (a)-(d)Results of structural matching for radical 阝 and the corresponding scores. (e)-(g) Results of structural matching for radical 阝 and the corresponding scores. —: the skeleton of the extracted radicals. - - : the skeleton of the input character.

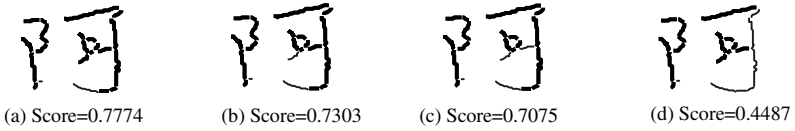


Fig.9. (a)-(d) Results of structural matching for character 阝 with the highest matching scores.—: the skeleton of the extracted character. - - : the skeleton of the input character. The final matching score is 0.7774.