

# An Evaluation System for String Extraction in the Airline Coupon Project

Yan Heping<sup>1</sup>, Zhiyan Wang<sup>1</sup>(Member,IEEE), Sen Guo<sup>1</sup>

<sup>1</sup> School of Computer Science of South China University of Technology,  
Guangzhou 510640, China

E-mail: gzyhp@126.com, wzhyan@ieee.org

## Abstract

*In this paper we introduce an evaluation system to evaluate the performance of segmentation algorithm aimed to the airline coupon project. It is based on two databases which are constructed by a standard observer method and three error measures. Due to the idea of consolidating subjectivity to objectivity, it is proved effective and objective in segmentation evaluation work in the airline coupon project. It is also convenient to analysis the segmentation results and to flesh out the databases endlessly in a net-based integrated working environment.*

## 1. Introduction

Our airline coupon project is about coupon recognition and management. Two essential problems are string extraction (by image segmentation) and character recognition. In the segmentation phase, string extraction is based on pixel level. It is done by classifying all pixels in an airline coupon image into two sets, one includes pixels belonged to the object characters we wish to extract from image, called object pixels; and the other includes pixels not belonged to the object characters, called background pixels. Due to segmentation results directly influence the precision of recognition, many efforts are made to extract the string objects more accurately in airline coupon.

Li's work[1] transferred to color space from the gray images and get better effect. But there are still many problems to resolve. First, we should have an objective method to evaluate the segmentation algorithms. In Li's work, he put forward two indicators to measure the segmentation efficiency, Rate of Removed Background Pixels and Rate of the Remaining String Pixels, but how to define string pixels and background pixels(called standard segmentation) is not mentioned. Second, we should have a way to analysis data of the segmentation images

to improve segmentation algorithm. Only the two indicators is not enough to do this.

According to the Ugly Ducking theory[2], evaluating is a considerable subjective process. So there is no standard segmentation in a pure objective way. By analysis the relationship between subjectivity and objectivity, we put forward a method called standard observer segmentation to define the two pixel sets in an objective way. The standard segmentation pixel sets and original images pair formed a dataset. Based on this evaluation database, three measure indicators are put forward to evaluate macroscopical and microcosmic performance of an algorithm.

Applying the standard observer method to form standard segmentation is not a easy work, for there always exists individual differences. We put forward a template method to resolve this problem. Character template are restored from samples by using the same standard observer method. All templates set a template dataset which are used to help standard observers to form standard segmentation.

There has been a limited amount of related previous work evaluating segmentation performance using datasets with human observers providing the ground truth. Heath et al.[3] evaluated output of different edge detectors on a subjective quantitative scale using criterion of ease of recognizability of objects(for human observers) in edge images. David Martin et al.[4] developed a database of human segmented natural images to evaluating segmentation algorithms. Closer to David's work is Sowerby image dataset that has been used by Huang[5] and Konishi et al.[6]. In our work, we construct two databases by standard observers method, provide an evaluation method, develop tools used for constructing the two databases interactively, and develop an evaluation frame model based on two databases and evaluation method. All of these have been integrated in a net-based integrated working environment(CoupEva 1.0). It is mainly used to evaluate segmentation algorithm and analysis data to

research new segmentation methods for airline coupon project.

This paper is organized as follow. In section 2, we describe the idea in detail of constructing the two databases. In section 3, we introduce the system, especially databases. In section 4, we define measures for evaluating segmentation and propose a frame model to evaluate. In section 5, we give an example that uses the system to evaluate performance of the algorithm in [1], and analysis results in detail. Then, we conclude in section 6.

## 2. The Idea for Constructing the Database

In a general way, segmentation algorithms are aimed to specific applications, so an evaluation method is also aimed to a specific application. For evaluation is a subjective process, so if segmentation results made little trouble to the following recognition process, we can say that the segmentation is efficacious. In the view of visual perception, we conclude that if there is little difference between segmentation image processed normally by perception and original image, we can say the segmentation is efficacious. But it is difficult to do this practically. And it is also difficult to measure the difference. For this is a subjective process as said before. We need to evaluate a segmentation algorithm in an objective and quantitative way. To do this we should consolidate the subjectivity to an objectivity.

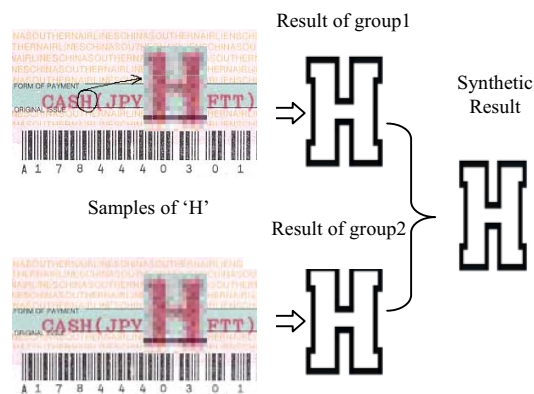
The problem can be considered in this way, a normal observer segments an image and produces a result image, we say the segmentation result is subjective. But when many normal observers segment the same image, result images will behave them with consistency in some degree. We consider this consistency represents a kind of objectivity in some degree; it obeys some statistical rules. In this way, we can consolidate subjectivity to objectivity. This idea is the foundation of our evaluation work.

In order to implement this idea, let a group of normal observers(called standard observers) to segment a selected sample image set and produce a set of result segmentation images. Because of consistency between their segmentation results, result segmentation images can be unified to a result. We call the unified segmentation result image standard segmentation image. Then we can use the standard segmentation image as a benchmark to evaluate segmentation algorithms. So the evaluation image database is composed of two parts, selected sample images and standard segmentation images.

In process of standard segmentation, we find that, by comparing segmentation result images of the same

original images produced by two standard observers, there is always a little difference, at least one or two pixels; even the same standard observer will produce the difference when he does segmentation twice. We think this is because of randomness in this process.

To our airline coupon images, we know that the printed characters should be composed of same pixels set, for they have been printed by same dot matrix printer. So we could use a set of dot matrix templates as a kind of standard character segmentation to help standard observers segment every character. In this way, we can get rid of the randomness. To do this, we should get dot matrix templates of every sample characters first. But most of them cannot be gotten directly. So, we used the same way as standard observers segmentation method to restore dot matrix templates from sample characters by two groups of standard observers, see Fig.1.



**Fig.1: The process of restoring the dot matrix by two groups of standard observers**

In the process of restoring dot matrix templates, we asked standard observers to obey the Gestalt's good form principles[7], which are called symmetry, tendency, and simplicity. For we think the design of dot matrix and human visual perception process obey these rules. In this way we construct a database of dot matrix templates which include all dot matrixes of characters appeared in sample images database. Its main application is to help standard observers to segment characters in sample images by using the dot template matching method.

In the standard segmentation process, we asked standard observers to use the corresponding character dot matrix template to match character in the original image, reserve pixels under template as object pixels, and clear other pixels(background pixels), even when the characters in the original image are in a distorting, skewing, pixels losing and et al case, see Fig.2. This is

an ultimate recognition rule consistent with human visual perception.

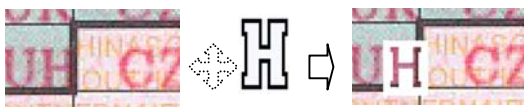


Fig.2: An example of standard segmentation

The process of segmentation for the character 'H', interactively, is followed by:

1. Pick the template 'H';
2. Move it to match the character image;
3. Adjust the position of template repeatedly, until they are in the best matching status in visual perception;
4. The system remains all the object pixels value, and set other pixels in blank, (255,255,255).

Note:

Just do the maximum matching, regardless character image skewing, object pixels losing, distorting and so on. For example, the object pixels in black form line are losing, but we should reserve them as object pixels.

Thus, the whole system includes two databases, one is an image database which is composed of original sample images and corresponding standard segmentation images, and the other is dot matrix template database.

## 2.1. Template Discussion

It is proved in practice that template method is useful to fulfill the idea of consolidating subjectivity and objectivity and eliminate randomness in consolidating process.

Though this method has obvious advantage in deal with small character set such as English characters, it also has obvious disadvantage. It is not suitable to deal with large character set, such as Chinese and so on.

Unlike method used by David Martin[4], this method is not a really ground-truth. For the standard segmentation process repaired the skewing, pixels losing, distorting and so on, see Fig.2. But it is more reasonable. Considering character recognition process of human visual perception, we can understand the ultimate recognition rule which ignored all of these. So we think it is consistency with the visual perception process.

## 3. The System and the Databases

### 3.1. The Construction of the Evaluation System

The CoupEva(ver1.0) system is composed of three subsystems, dot matrix restore subsystem include dot matrix database, standard segmentation subsystem include image database which composed of original sample images and standard segmentation images, and performance evaluation subsystem. The first two subsystem including a working database both work in the same way. They are interactive environments related to subjective work. The system construction is shown in Fig. 3.

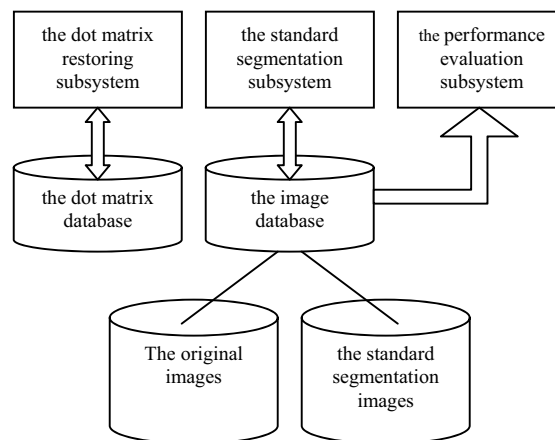


Fig.3: The construction of the evaluation system

### 3.2. The Method of Databases Constructing

To construct the two databases, we first select two groups of standard observers, and a dataset of original sample images. Before actual working, these two groups of standard observers received training for a period of time, which include project construction, gaining acquaintance with dot matrix, and some practice lessons. Then the first construction phase is to construct dot matrix database. In this phase, standard observers restore all of dot matrix templates of characters appeared in original sample images. All these templates composed dot matrix template database. When this database formed basically, it is the time to construct image database by using it. In the image database construction phase, all of the original images have been segmented by using dot template matching method. The pairs of original sample images and standard segmentation images build the image database.

In the phase two, construct the two databases, all of the characters whose dot matrix template to be restored or the original images to be segmented are assigned to all members in the group randomly. Then all of them

finish their work. Their products build a dataset. The other group also finishes the same work independently. Then, the two result datasets are compared to find their differences one by one. Full discusses are hold to analysis the difference until get their act together. Up to now, the construction work of the two databases is accomplished. But the databases are not unchanged forever. The databases should be expanded and the elements in the two databases should be perfected in the future application, whenever new events happened.

### 3.3. The Status of the Databases

Now, the dot matrix database includes the dot matrix templates of 41 characters appeared in the selected original sample images. There are two kinds of font appeared in the sample images, so each character has two counterpart of dot matrix templates. Now, the dot matrix database includes 82 dot matrix templates totally.

The original images in the image database include five kinds of original sample images which is selected from three airline companies in China and grouped by visual quality, each kind have 20 samples and all are in  $1126 \times 502 \times 24$ bits, uncompressed RGB format. This image database also include four kinds of sub-images of the original images for special analysis usage, each kind have 10 samples; they also in 24bits uncompressed RGB format, but not in same size. In the standard segmentation images, object pixels are kept the same value with the corresponding pixel value in their original images, and background pixels are set in a fixed value, (255,255,255).

## 4. Segmentation Error Measures and the Evaluation Method

In order to evaluate the performance of a segmentation algorithm in the independent segmentation phase, we need to compare result images segmented by the algorithm and the standard segmentation result images, one pixel by one pixel. We use three segmentation error measures as quantification indicators.

For a given original sample image,  $T$  represent pixels set. Pixels in the image should be classified into two class, the object pixels class and background pixels class. Let  $F$  represent the object pixel set, and  $B$  represent the background pixel set, certainly we have:

$$T = F \cup B \quad \text{and} \quad F \cap B = \emptyset$$

Let  $F_s$  and  $B_s$  represent the object pixel set and the background pixel set of the standard segmentation image; and let  $F_a$  and  $B_a$  represent the object pixel set

and the background pixel set of the algorithm segmentation image. We also have:

$$T = F_s \cup B_s \quad \text{and} \quad F_s \cap B_s = \emptyset$$

$$T = F_a \cup B_a \quad \text{and} \quad F_a \cap B_a = \emptyset$$

By compare each pixel in two object pixel sets and two background pixel sets, we can compute two counts of pixels that are classified by mistake. Let  $C_b$  represent counts of pixels which should be classified to background pixel set but have been classified to object pixel set, let  $C_f$  represent counts of pixels which should be classified to object pixels set but have been classified to background pixels set. Let  $|X|$  represent the counts of set  $X$ . Note that:

$$C_b \neq |B_s| - |B_a| \quad \text{and} \quad C_f \neq |F_s| - |F_a|$$

They are counts of different pixels between two sets by comparing the two sets one pixel by one pixel.

Then we define three error measure indicators:

$$\mu_f = C_f / |F_s| \quad (1)$$

$$\mu_b = C_b / |B_s| \quad (2)$$

$$\mu_t = [C_f + C_b] / |T| \quad (3)$$

$\mu_f$  is the object pixel misclassified rate,  $\mu_b$  is the background pixel misclassified rate, and  $\mu_t$  is the total misclassified rate. The two indicators  $\mu_f$  and  $\mu_b$  are mainly used to analysis the algorithm and data in detail. The indicator  $\mu_t$  is mainly used to evaluate the integrated performance of the algorithm. Obviously:

$$\mu_t \neq \mu_f + \mu_b$$

The evaluation frame model based on the three indicators is shown in Fig.4, which is in the 3rd subsystem of the evaluation system. When we evaluate an algorithm use this frame model, all original images in the image database are segmented by this algorithm and produce corresponding result segmentation images, each include object pixel set and background pixel set. This evaluation subsystem counts the means of three

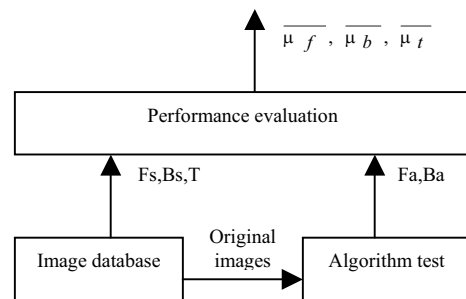


Fig.4: The frame model to evaluate

error measure indicators to evaluate performance of the algorithm.

## 5. Use the Evaluation System to Evaluate Segmentation Algorithm

In this section, we use our system to evaluate and analysis a segmentation algorithm. To analysis the algorithm in [1], we made a post-process to the result segmentation images segmented by this algorithm. In post-process, we get rid of detached noise pixels by applying a filter to these result segmentation images. Tab.1 is the results:

**Tab.1: Error measure indicators to algorithm[1]**

	$\mu_b$	$\mu_f$	$\mu_t$
Have no post-process	0.1362	0.0459	0.1316
Have post-process	0.0931	0.0486	0.0908

The evaluation results are almost the same as results in [1]. The  $\mu_b$  is much better after the post-process. This is because of that most of the pixels misclassified to object pixel set are gotten rid of by the post-process. To the contrary, the  $\mu_f$  is a little worse after the post-process. Because a few of object pixels are filtered by the post-process; but we can see that the amplitude is minor and so acceptable. This is also a method to evaluate the performance of the post-process algorithm.

But it is not enough to evaluate the algorithm by the two indicators. We can think about two extreme cases, one is that all of the pixels in the original image are classified to the object pixel set, and the other is contrariwise. In the first instance, the  $\mu_f$  is the best, 0; and the  $\mu_b$  is the worst, 100%. So, only the two indicators cannot evaluate integral performance of the algorithm.

We can see in Tab.1,  $\mu_t$  is brought down by 4 points; that indicate the post-process is effective. All three indicators can perform an integral and pinpoint evaluation.

## 6. Summary and Conclusion

In this paper, we introduced an evaluation system which include two databases. This system have two main functions; one is to flesh out the two database constantly, and make our evaluation more and more

objective; and the other one is to evaluate a segmentation algorithm, it is also the main function of this system. The two databases is constructed by a method that consistency with human visual perception, and possess a measure of statistical objectivity. This system is mainly aimed at the airline coupon project, but not just limited to this project. It can be extensively used to other comparable applications, just not in large character set.

## Acknowledgments

We would like to thank the students of SCUT CT01, especially the two group leaders, Huang Jiahui and Liu Xiong, who made many efforts to construct our databases.

This research work is supported by GuangDong Provincial Natural Science Funding Project B6-109-497,China.

## References

- [1] Yi Li, et al, "String Extraction in Complex Coupon Environment Using Statistical Approach", *Proc of 7<sup>th</sup> International Conf. on Document Analysis and Recognition*, 2003, Edinburgh, pp. 289-294.
- [2] M.W.Eysenck,M.T.Keane, *Cognitive Psychology: A Student's Handbook*, Fourth Edition, Teacher's College of East of Chain Press, XiangHai, 2003.
- [3] M. D. Heath, S. Sarkar, T. Sanocki, and K.W. Bowyer, "A robust visual method for assessing the relative performance of edge-detection algorithms", *PAMI*, 19(12), pp. 1338-1359, 1997.
- [4] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms", *ICCV 2001*, vol.2, pp. 416 - 423
- [5] J. Huang, *Statistics of Natural Images and Models*, PhD thesis, Brown University, May 2000.
- [6] S. Konishi, A. L. Yuille, J. Coughlan, and S. C. Zhu, "Fundamental bounds on edge detection: an information theoretic evaluation of different edge cues", *CVPR*, pp. 573-579, 1999.
- [7] Kurt Koffka, *Principle of Gestalt Psychology*, ZheJiang Education Press, HangZhou, 1997.