2003 International Symposium on Advanced Intelligent Systems
September 25-28, 2003, Jeju, Korea

Knowledge Based Intelligent Photoshot-to-Translation System

Tam Heng Wa
Electrical and Electronics Engineering
FST, University of Macau
Address: Rotunda Da Maratona, BL. 04, FL. 07, FLAT O
ED. MEI KENG GDN-MEI FOK KOK, Macau, China
email: m986241@wkg1.umac.mo

Abstract - In recent years, most of the researches on pattern recognition are for medical diagnosis or for characters recognition. In fact its applications are very wide. In this paper, the pattern recognition is employed by linguistic translation, i.e. the output of Pattern Recognition is translated into another language. In this paper, it focuses on several fields: (1) System overview—explicate the functions of each part individually; (2) Criteria on the system—discuss the difficulties in each part; (3) System implementation—discuss how to design the approaches for constructing the system. Furthermore, intelligent approaches are considered be use on the system in different parts. They are discussed in the late paper, and also we concentrate on user interface, which can make a serious of processes in order, and easy control—just only pressing a few buttons. It is a new and creative attempt in digital system.

Keywords: pattern recognition, Optical Character Recognition (OCR), Linguistics Translation

I. INTRODUCTION

In recent years, most of the researches on pattern recognition are for medical diagnosis or for characters recognition. In this paper, the pattern recognition is employed by linguistic translation. Furthermore, an integrated user interface is set up to coordinate these two parts.

In general a digital camera is a device to implement the image data handling, it converts the analog video originated from the object to digital image, and store the image for the future utility. In this processing, there are three primary image data handling functions—image digitization, image storage and image display, eventually digital images are produced. A digital image is simply a matrix where each number represents the brightness at regularly spaced points or very small regions in the image. These call pixels (picture elements) and the brightness value of a pixel is called its gray level [1]. The elementary function is implemented by Image digitization, and in this process it can be divided into two principal functions, image sampling and image quantization. The latter is also referred to as analog-to-digital conversion. After image digitization, the next step is image storage, which provides long-term storage of image data for processing, display and archive purposes. Finally image display converts a digital image back to the analog electrical form on a video display monitor [7].

In the following, it introduces Optical Character Recognition (OCR). OCR is one of pattern recognition for image data processing and it can extract the character from digital images. The common approach is Statistical pattern recognition, which assumes that the images may contain one or more objects and that each object owning its property of type, category, and pattern class [2]. OCR contains three phrases: image segmentation, feature extraction and classification.

After OCR, its output is transported to the linguistic translation, which is replacement of a representation of a text in one language by a representation of an equivalent text in a second language [4]. In the input pattern its objects (words) in the text should have been classified, then they are extracted from digital images, and translated into another language. Therefore, except the domain of linguistic database, the accuracy of the translation also depends on the capability of OCR. Moreover there are many different formats of digital images produced by digital cameras or web cams, and these images contain distinct types of fonts, the textures of the object, and symbols or graphics, even anomaly characters and noisy characters caused by image processing. This situation not only occurs in the simple black-white photos and low color resolution photos, and including high color resolution photos.

II. DESIGN OF THE SYSTEM

A. Overview of The System

Before going to design of the system, we should define our scope of research:

(1) Production of digital images
(2) Image preprocessing
(3) Optical Character Recognition
(4) Image post processing
(5) Translation from English into Chinese

After determine its steps, its system diagram can be obtained as figure 1.

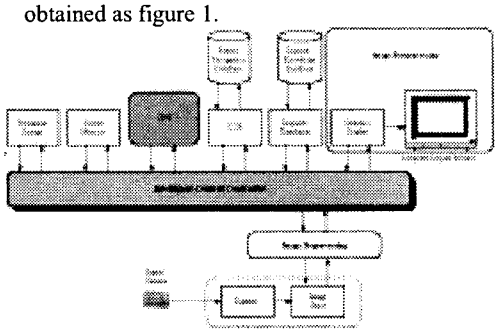

Figure 1 System Diagram

B. Architecture of The System

(1) Production of digital images

It is just the transformation from optical signal to electrical signal by using a sensing device. The analog (electrical) signal is transformed to a digital one by a video digitizer (frame grabber) [5], and after being kept in image store, the image goes into the process of image data processing.

(2) Image preprocessing

Since each type of OCR approaches also exists his constraints such as image format, color resolution and font types. At first a digital image must be judged if it can be analyzed in the OCR process. If the digital image is out of range, it should be ignored. Or if the image is close to the boundary, then the system should try to convert the image into an optimal state such as adjusting its gray level, and decrease the size or color resolution in the images.

(3) Optical Character Recognition

OCR contains three phases: image segmentation, feature extraction and classification.

In image segmentation, or object isolation, each object is found and its image is isolated from the rest of the scene. In this phase, there are four processes, region approach, boundary approach, edge approach and binary image processing [2]. (a) In region approach, it provides gray level thresholding can filter the object unwanted. (b) Boundary approach is gradient-based segmentation to find the edges directly by the high gradient magnitudes. (c) Edge approach is edge detection and labeling. These two latter approaches can fix the complexity of text and pictures. And also in edge approach, Hough Transform can be used to solve the uncompleted character, caused by noise. (d) Binary image processing, binary images—those having only two gray levels—constitute an important subset of digital images. If the initial segmentation is not completely satisfactory, it can be improved in binary image processing [2].

In feature extraction, the objects are measured. A measurement is the value of some quantifiable property of an object. A feature is a function of one or more measurements, computed so that it quantifies some significant characteristic of the object. The feature extraction process produces a set of features that, taken together, comprise the feature vector. This drastically reduced amount of information (compared to the original image) represents all the knowledge upon which the subsequent classification decisions must be based. It is productive to conceptualize an n-dimensional space in which all possible n-elements feature vectors reside, thus, any particular object corresponding to a point in feature space.

In the last phase, classification process, its output is merely a decision regarding the class to which each object belongs. Each object is recognized as being of one particular type, and the recognition is implemented as a classification process. Each object is assigned to one of several pre-established classes that represent all the possible types of objects expected to exist in this image [2].

Although the above discussion only on black and white image, it is also available on color image, because color image may be thought of as a combination of three independent gray-level images. Color image analysis do not differ significantly from those applied to black and white image; they just entail application of the same methods as those for black and write image, but applied threefold to the different color images [6]. The IDEF0 diagram of the system can refer to Appendix A.

(4) Image post processing

After the characters are recognized, the alphabets will be printed in the output window, and saved in an editable text file, so it is ready for the next process, linguistic translation.

(5) Linguistics Translation

In general, the basic processes of translation are: preprocessing, analysis, transfer, synthesis, and post processing. Within analysis and synthesis, there are three distinguishable areas of operation: syntactic, semantic and pragmatic. We can see the translation processes in figure 2. However, the process is not a linear one in which stage follows stage in a strict order. It is an integrated process in which, although every stage must be passed through, the order is not fixed and back-tracking, revision and cancellation of previous decisions are the norm rather than exception. [4]

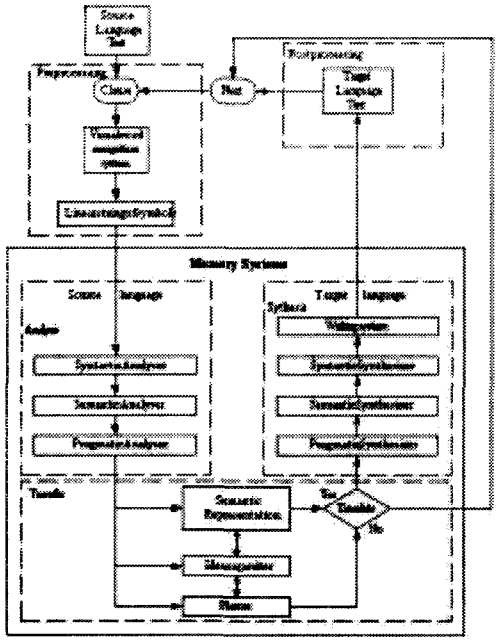

Figure 2 Translation Process: Outline Model

III. CRITERIA ON THE SYSTEM

A. Easy Control Man-to-Machine Interface

To make OCR and Translation co-operating each other, an integrated platform is being used. It can make the system to call the different threads in sequence: image preprocessing, pattern recognition, image post processing, and translation. Furthermore the platform should obtain one-time setting characteristics: set the configuration on the first time running, then repeat the steps memorized automatically. Therefore it can reduce the repetitive times for pressing the same buttons. After the integration of all processes being set up, the system can be programmed to operate in sequence. Eventually it can be controlled in a simple way—just press a few keys to run the system after first time running.

B. Optical Character Recognition

In the part of OCR, although image segmentation can remove most of noise and extract the objects from their background, there is a lot of work to do before identifying the objects.

First step is size measurement, contains area and perimeter measurement, especially for polygon measurement, and objects with curvature, it requires a lot of numerical calculation and time. Then shape analysis is to classify shape of objects, such as rectangularity, circularity etc physical dimensional measures. Sometimes if the shape is irregular, Invariant moments and shape descriptors will be used to measure the object's characterizations.

On the other hand, if the spatial frequencies of an object in a picture are low, its brightness has small variances, so we can suppose the object with smooth surface. Or if the spatial frequencies are high, its variances in brightness are large, its surface is supposed as coarse. So it is common that there is some texture on the objects, so texture analysis also is necessary. Final step is curve and surface fitting, and its functions are guessing the original shape and feature of object, and also surface fitting can be used for noise removal and for purpose of measurement [2].

After above steps, the alphabets in images can be extracted. Otherwise, the process of OCR is unidirectional, have no cycles, so Feedforward Neural Network is suggested to be used, because the network is suitable to forward chaining cases.

C. Knowledge based Translation System

What is knowledge based system? A knowledge based system can be defined to be:

A computerized system that uses knowledge about some domain to arrive at a solution to a problem from that domain. This solution is essentially the same as that concluded by person knowledge about the domain of the problem when confronted with the same problem. [8]

In general knowledge based system can be implemented in two major architectures: knowledge base and inference mechanism. These architectures define how the rules are expressed and related to the facts. The knowledge base consists of predicate calculus facts and about the subject at hand. The inference mechanism consists of all the processes that the knowledge base to deduce information requested by the user—resolution or forward or backward chaining [9].

Furthermore, on translation, knowledge base and inference mechanism have their definitions [4]:

(1) **a knowledge base** consist of :
   (a) source language knowledge; the syntactic rule systems of the code, its lexicon and semantics and its text-creating systems
   (b) target language knowledge; equivalent to that in the source language
   (c) text-type knowledge
   (d) domain knowledge
   (e) contrastive knowledge of each the above;

(2) **an inference mechanism** which permits:
   (a) the decoding of texts, i.e. reading and comprehending source language texts
   (b) the encoding of texts, i.e. writing target language texts e.g. a writer's assistant system which helps with writing.

The research and development of translation process can be refereed to the paper "Machine Intelligence from Vision to Translation" [3], Mr. Dong Qing fu, Tsinghua University.

IV. IMPLEMENTATION OF THE SYSTEM

To achieve only pressing a few buttons, it is necessary to develop a platform to integrate photo capture, pattern recognition and translation, and make them coordinate in sequence. The first thing to do is build up a flow diagram as appendix B, then write the program to call different sub-program to accomplish different task, and arrange them in order. Finally it is to adjust the timing between each part.

However, its database is not omitted, it stores all data in each part, like storage of photo, the order of processes, exemplars, rules of each parts and texture analysis, histograms and the knowledge base and inference mechanism of translation. The database is a fundamental element for intelligent system.

In the part of OCR, after having got the measurement of the object, comparing measures with known features is performed. However, since the size and the shape of objects may be changed after a series of image processing. It is possible that misclassification can be effected by the assignment of an object to an inappropriate class. It has to estimate by the similarity of known features. But the rules of decision must be accordingly to the expert knowledge. Therefore knowledge database is a wide comprehension on each part like pattern recognition and it stores the types of fonts, categories and color of each object, it can make the system easy to find out the object's feature vector, and shorten the time for classification.

V. CONCLUSION

As we discussed in the preceding, Knoledge Based Intelligent Photoshot-to-Translation System can be concluded into three main research scopes:

1. Easy man-to-machine control
2. Feedforward neural network application on OCR
3. knowledge based linguistic translation

In these three directions, the previous is essential for constructing a user friendly interface, its goal is to make the simplest and effective approach—just pressing as few keys as possible to run the system; Although OCR is straight forwarding, the more complicate the network, the more layers in the network. If there are too many layers, it is possible that over-fitting occurs, so its estimation on the out-of-sample error for validation is interesting; Finally, because the linguistic translation and knowledge based system are two wide topics, they are very important and complex in the whole system, especially the rule base in knowledge based system is close relating to translation, so it is worthwhile to set up a complete and effective database in the future.

REFERENCE

[1] Earl Gose, Richard Johnsonbaugh and Steve Jost, Pattern Recognition and Image Analysis. Prentice Hall PTR, 1996.

[2] Kenneth R. Castleman, Digital Image Processing, Prentice Hall, Tsinghua University, 1998.

[3] Dong Qing fu, Machine Intelligence from Vision to Translation, Tsinghua University, 2002.

[4] Roger T. Bell, Translation and Translating: Theory and Practice, Foreign Language Teaching and Research Press, 2001

[5] I Pitas, Digital Image Processing Algorithms and Applications, John Wiley & Sons, INC, 2000.

[6] Michael Seul, Lawrence O'Gorman and Michael J. Sammon, Practical Algorithms for Image Analysis Description, Examples, and Code, Cambridge University Press, 2000.

[7] Gregory A. Baxes, Digital Images Processing, Principles and Applications, John Wiley & Sons, Inc., 1994

[8] Avelino J. Gonzalez & Douglas D. Dankel, The Engineering of Knowledge-Based Systems Theory and Practice, Prentice-Hall, 1993.

[9] Nils J. Nilsson, Artificial Intelligence, A New Synthesis, Morgan Kaufmann Publishers, 1998.