MVA'90 IAPR Workshop on Machine Vision Applications Nov. 28-30,1990, Tokyo

Development of an Automatic Recognition System for Plant Diagrams

Takashi Futatsumata, Go Shichino, Jun'ichi Shibayama & Akira Maeda

Information Systems & Electronics Development Laboratory,
Mitsubishi Electric Corporation.
5-1-1, Oofuna, Kamakura City, 247 Japan

Abstract

In electric generation plants, facility administration is going to be computerized for secure and efficient maintenance, and it is needed to input many already existing plant diagrams (so called piping & instrument diagrams) into computer systems. Then it is required to recognize plant diagrams, whose drawing quality is so tough for recognition, that the drawn elements, namely, symbols, lines and characters, often have contacts with each other.

Thus we developed an automatic recognition system for plant diagrams, which enabled to input such diagrams at recognition rate over 80 %. And together with an efficient verification process, the total time to input a diagram is reduced to be under 30 %, compared to conventional CAD-system input.

1. Introduction

Facility administration of electric generation and other plants, such as maintenance and overhauling of instruments and pipings, are based on plant diagrams. In order to computerize this administration, it is needed to input many already existing plant diagrams into computers. It is required to recognize the diagrams automatically, since a conventional CAD-system input needs much cost [1].

There have been attempts to develop automatic recognition systems, but most of them did not treat diagrams with much complexity. Our development aims to input such type of diagrams.

2. System Overview

2.1 Features of the system

The plant diagrams to be recognized contain many elements, and as they are not intended for computer input, they have many contacts of symbols, lines and characters with each other. And the diagrams are in poor drawing quality due to reproduction with size reduction and retouching correction (Fig. 1).

Most of the symbols consist of *a basic figure* (as a pair of triangles in valve symbols) and some *additional figures* (Fig. 2), and there are hundreds of their classes.

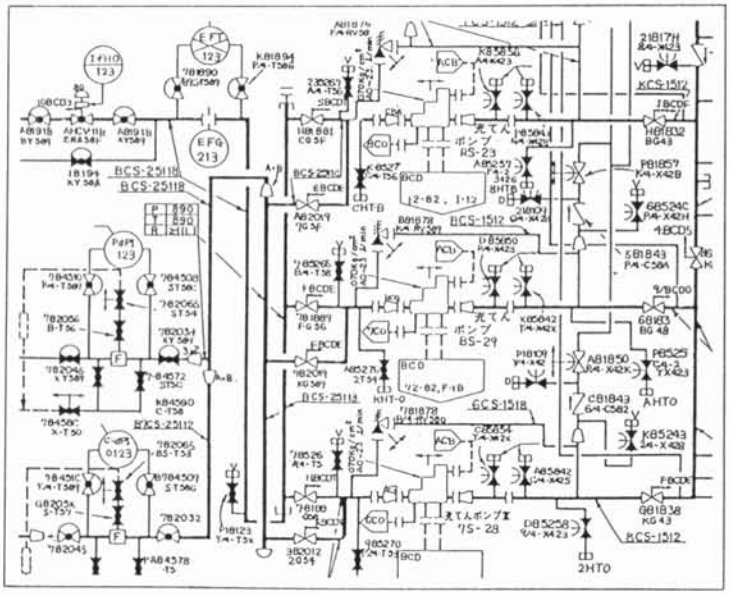

Fig. 1 An example of a plant diagram

Fig. 2 Samples of symbols

Thus we developed a system to input such diagrams efficiently with a combination of recognition and verification process, with the following features.

(1) *Recognition;* The items, which need much correction time, such as the positions of symbols, lines and characters, are to have high recognition rate. But the items, which need less correction time, such as the classes of symbols and characters, may have not so high rate.

(2) *Verification;* For the classes of symbols and characters, whose recognition rate tends to be not so high, verification and correction are to be done efficiently.

2.2 Chief specifications of recognition

The following items are to be recognized.

(1)*symbols,* (i.e. valves and other instruments); their position, orientation and symbol class.

(2)*lines,* (i.e. pipings and measurement/controlling wires); their position and connection information to the symbols.

(3)*character strings,* (i.e. attributes of symbols); their position and character codes.

2.3 System configuration

The developed recognition system consists of an image reader, a minicomputer and a workstation (Fig. 3). The minicomputer is equipped with an image processor and a character recognition processor.

The minicomputer recognizes the plant diagrams input by the image reader, and the workstation is for verification and correction of the recognized data. The data are finally converted to the format required by the facility administration system.

3. Recognition Process

The recognition process of the plant diagram is as follows (Fig. 4).

3.1 Classification of the elements

After the vectorization of the raster input image, the drawn elements are roughly classified. The elements in contact with each other are not dealt here. (See 3.3).

(1) *Detection of characters*

A block of connected vector elements is measured, and a block with height and width less than thresholds is considered to be a character candidate.

(2) *Detection of connection lines*

From the vectors (without the character candidates), a horizontal or vertical vector with length more than a threshold is detected to be a line candidate. And a vector in its elongation, or a vector perpendicular to it, is also judged to be a line candidate. The other vectors become symbol candidates.

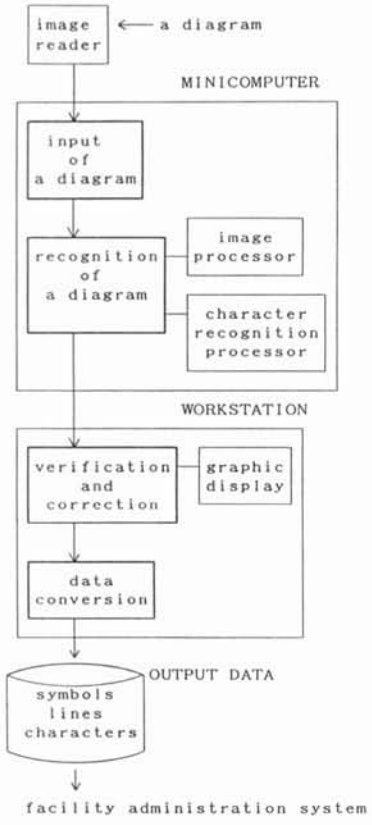

Fig. 3 System configuration

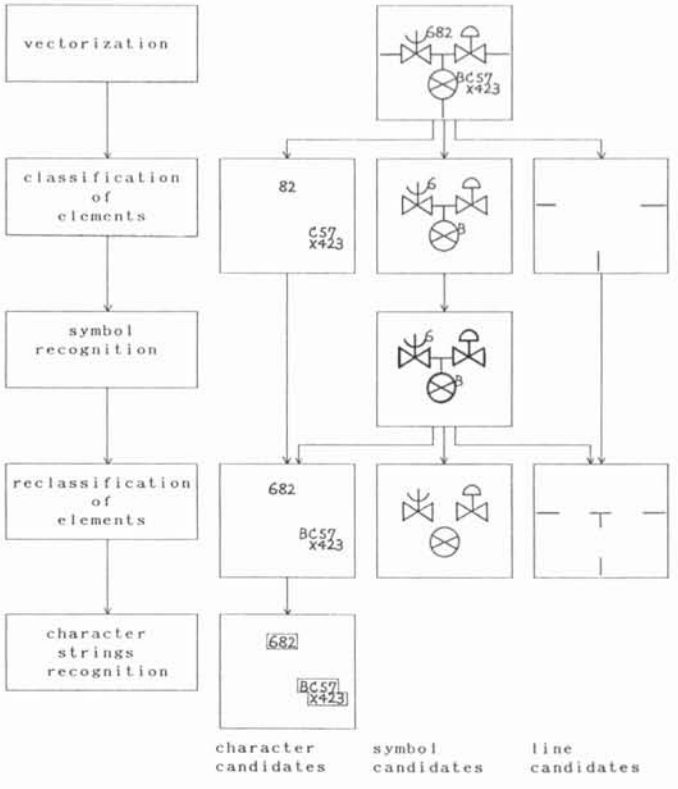

Fig. 4 Recognition flow

3.2 Symbol recognition

From the symbol candidates, characteristic figures such as loops or black regions are detected, and they are matched to the symbol dictionaries, so that the detection and recognition of the symbols are carried out parallelly. For detail, see the Section 4.

3.3 Reclassification of the elements

The symbol candidates contain characters and short connection lines, drawn in contact with symbols. Thus, the vectors not detected as symbols in 3.2 are reclassified to be character or line candidates. For this process, utilized are the recognition results of the symbols, such as;

(1) connecting position to a symbol,

(2) rules, for example, that a vector between 2 symbols, is a connenction line.

A line candidate vector and informations, as to which symbols the vector is connected, together make up connection data.

3.4 Recognition of character strings

From the character candidates, a set of vectors within a threshold height are detected to be character strings. 2 character strings may be touching each other vertically, so a string over a threshold height is divided into 2 strings [2].

4. Symbol recognition

4.1 Features of the symbol recognition

The features are to treat a plenty of symbol classes, and to cope with contacts with other elements.

First detected are *basic figures*, stable to contacts, and common to plural symbol classes. Then the area around *a basic figure* is analyzed to detect *additional figures*. The combination of *a basic figure* and some *additional figures* decide the final symbol class. In detecting *additional figures*, segmentation is recursively called to detect additional figures in contact with characters.

4.2 Flow of recognition

The symbol recognition is carried out as follows (Fig. 5).

(1) *Detection of basic figures*

From the symbol candidates, detected are basic figures such as closed loops, black regions, T-type figures etc, of certain sizes.

(2) *Detection of additional figures*

The area around a basic figure is examined, and additional figures are detected and classified (See 4.3). They are symbol constituents other than the basic figures, such as, rods, arcs, small rectangles, etc., describing detail class of valves.

(3) *Classification of a symbol*

A symbol is classified totally by the basic figure, additional figures around it, their connections, positions, etc.

4.3 Recursive segmentation

In 4.2(2), additional figures are detected and classified simultaneously, to acquire symbols in contact with characters. Namely, the area around a basic figure is segmented and matched to the dictionary of additional figures. If this fails, then another segmentation is recursively called, and matching is re-tried. This is to get rid of errorneous segmentation due to the characters in contact with symbols (Fig. 6).

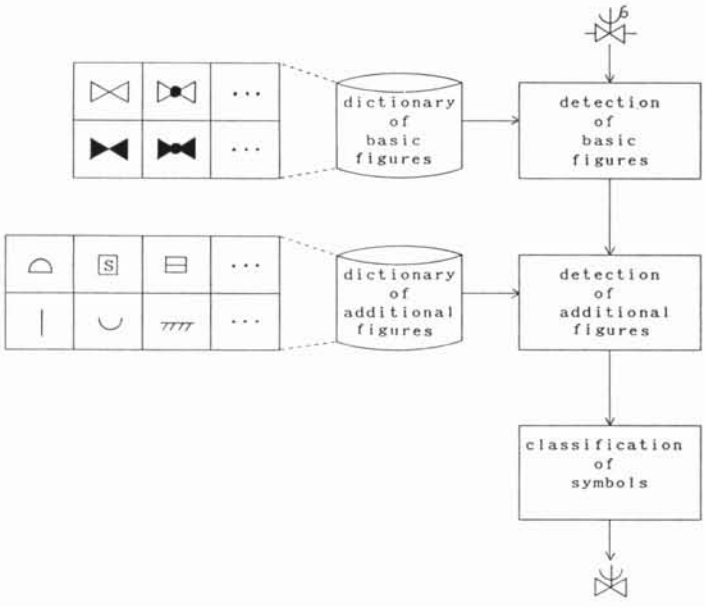

Fig. 5 Symbol recognition

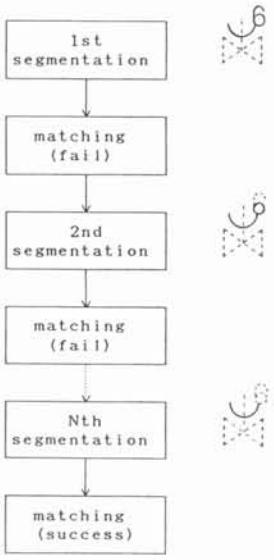

Fig. 6 Recursive segmentation

5. Verification Process

Recognition errors of the character and symbol classes can be verified and corrected in the following process [4].

5.1 Verification of character strings

The raster image of the detected character strings and their recognition results are shown parallelly at a work-station's CRT display, for an operator to verify and correct the characters recognized. By this method, the operator does not need to verify the results referring directly to the complicated plant diagram, thus enabling efficient verification (Fig 7).

5.2 Verification of symbols

Hundreds of the symbol classes have been sorted by the basic figures and assigned hierarchically to verification menues. When an operator picks a misclassified symbol in the CRT display, a menue with the same basic element appears automatically, to allow the operator to select the just symbol. The probability, that the error symbol and the just symbol are in the same menue, is high, since the basic element hardly get misclassified (Fig. 8)

6. Conclusion

For plant diagrams, whose quality of drawing is tough for recognition, we developed an automatic recognition system. The main results of the development are as follows;

(1) Adoption of sophisticated methods enabled to recognize such diagrams at recognition rate over 80 %.

(2) Together with an efficient verification process, the total time to input a diagram is reduced to be under 30 %, compared to conventional CAD-system input.

References

[1] H. Harada et al., "Recognition of Freehand Drawings in Chemical Plant Engineering", *IEEE Computer Society Workshop on CAPAIDM*, pp. 146 - 153, 1985.

[2] Go Shichino et. al., "Character Strings Segmentation for Automatic Input of Plant Diagrams", *Proc. of 41st Annual Convention IPS Japan*, 5M-9, p. 2-215, 1990 (in Japanese).

[3] A. Maeda et al., "A Recognition Method for Network Diagrams", *International Workshop on Machine Vision and Machine Intelligence*, pp 208 - 213, 1987.

[4] J. Shibayama et al., "Interactive Verification Process for Automatic Input of Plant Diagrams", *Proc. of 39th Annual Convention IPS Japan*, 3E-3, p. 504, 1989 (in Japanese).

Fig. 7 Verification of character strings

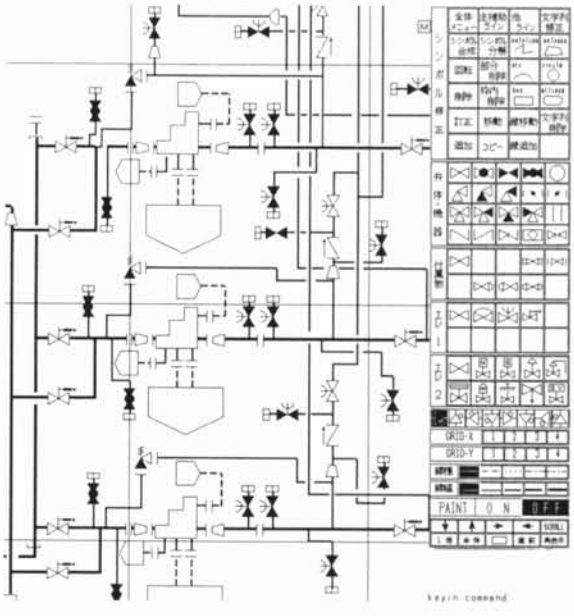

Fig. 8 Verification of symbols