Development of a system for practical skill training of maintenance personnel

Yu Yating 1, Yuji Yamasaki 1, Yasuhiro Kajihara 1, Tomoyuki Akashika 2, Kazuhiko Izumi 3, and Mitsuru Jindai 4

1 Graduate School of System Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, Tokyo, 191-0065, Japan
2 Sanky Inc., 5-23 Kachidoki 6-Chome, Chuo-ku, Tokyo 104-0054, Japan
3 infoSense Corp., 5-23 Kachidoki 6-Chome, Chuo-ku, Tokyo 104-0054, Japan
4 Graduate School of Science and Engineering, University of Toyama, 3190 Gofuku, Toyama, Toyama, 930-855, Japan

Abstract: Japanese companies offering plant maintenance services often employ workers from Asian neighbors. Those workers receive practical skill training at training centers before being dispatched to the various plants where they actually work. This paper proposes a method and a system for providing those workers with the practical skill training. The system judges progress in the work from the images of the facilities to be maintained and gives step-by-step instructions to the workers in accordance with the progress in the work. The system employs the client-server architecture of processing. A client displays operation manuals to the workers nearest to the facilities. The server judges the progress in the maintenance work and combines all the data for the practical skill training. In this paper, the validity of the method is verified by a trial operation and discussed from the standpoint of the processing time and the determination of the progress in the work.

Key Words: Plant maintenance, Industrial engineering, Image processing, Practical skill training

1. Introduction

Chemical plants are designed for long-term use and the maintenance of their facilities is important to keep the plants in a good condition. However, due to increasing long-term total costs, the maintenance work has often been outsourced. As a consequence, the market size of the long-term contracts of the facility maintenance work has been constantly growing[1]. The maintenance workers employed by Japanese companies are often from Asian neighbors. After receiving the training for the maintenance of the facilities at training centers in Japan, they are sent to the plants in Japan or the Middle East for conducting the maintenance work. The regulatory restrictions in Japan, however, has curtailed the period of new employees’ internship and moreover, the use of the facilities has been diversified. These circumstances make it hard for the training centers to provide the workers with sufficient work training for the various facilities. As a result, new employees are often sent to their work sites without receiving sufficient training. In light of these problems, we developed a prototype of a practical skill training system[2], which is designed to be used at the work sites both in Japan and overseas and can be installed in tablet PCs. Pictures of each special tool are taken by a camera mounted on the back of a tablet. Image processing is performed on the shotted images. The kind of a tool is identified by template matching. Following the identification of the tool, a tutorial manual for the usage of the tool is displayed in video, pictures and sentences on the display of the tablet.

This paper aims further at the development of our prototype to provide workers with an innovative function that can judge the progress in the work from pictures taken of the facilities. This function allows step-by-step instructions to be given to the workers in accordance with the progress in the work. Additionally, the configuration of the system is changed from a stand-alone type to a client server. The client shows operation manuals to a worker and the server judges the progress in the maintenance work and stores information concerning the operation manuals of all the facilities. Data transmission between the two components is carried out over the Internet. This configuration allows the server to gather all the data for the practical skill training.

2. Proposed Method

2.1 Outline of the practical skill training system

This system uses image processing to automatically recognize each of the special tools and judges the progress in the work from the pictures taken of the facilities. Fig.1 shows a scene in which two workers use this system to recognize a tool laid on the floor. Template matching is used in this image processing. A proper operation manual for the practical skill training is chosen and displayed on the monitor of a tablet in accordance with the result of template matching. Fig. 2 shows an example of an operation manual. The work procedure is summarized in a flow chart on the right-hand side and detailed explanation is shown in a text, pictures and videos in the center. The pictures of the main tools are shown on the left-hand side. If you click on a picture of a tool, you will be guided to a page which offers detailed explanation of the tool.

Our prototype system uses SSD (Sum of Squared Difference)
templates. Consequently, we decided to apply SURF (Speed Up Robust Features) [3] to template matching for the curtailing of the processing time.

The procedure for the judgment of the progress in the work is shown below.

### 2.2 Judgment of the progress in the maintenance work

The entire operations of the maintenance work are composed of a plurality of tasks. The template image of every task is prepared in advance. Pictures of a facility are taken as input images as the work progresses. Template matching by SURF [3] is performed on each of the input images. SURF detects the key points of both images. An example is shown in Fig. 3. Some key points were detected in the background of a facility, i.e. key points of lines on the floor. Accordingly, we introduced a step to exclude the key points that deteriorates the performance of the template matching.

In connection with all the detected key points, the step checks a gap in the coordinates of an input image and a template image. Subsequently, it excludes the key points having the large gap as non-coincident key points. Greater details of this procedure are given below and incorporated in SURF for the improvement of the accuracy of the template matching.

#### 2.2.1 Exclusion of the non-coincident key points

SURF detects sets of key points having the same characteristic quantities both in the input and template images. Let represent the coordinates of those key points as:

- **Key points in an input image**
  \[ p_k = (i_k, j_k), \quad (k = 1, 2, \cdots, n) \]

- **Key points in a template image**
  \[ g_k = (u_k, v_k) \]

A gap between the k-th key points of both images is represented by:

\[
L_k = \sqrt{(i_k - u_k)^2 + (j_k - v_k)^2}
\]

An arithmetical mean of \(L_k\) is given by:

\[
L_{ave} = \frac{\sum_{k=1}^{n} L_k}{n}
\]

A standard deviation of \(L_k\) is given by:

\[
\sigma = \sqrt{\frac{1}{n} \sum_{k=1}^{n} (L_k - L_{ave})^2}
\]

Next we set criteria to exclude non-coincident key points based on the statistical hypothesis testing approach.

\[
R = \{ p_k | (L_k - L_{ave}) > \alpha \cdot \sigma \}
\]

All the key points included in the set \(R\) are excluded and the coefficient \(\alpha\) is deemed to be determined in accordance with the conditions of an actual work site. The bigger the coefficient \(\alpha\) over 1.0, the greater the risk of the adoption of the non-coincident key points. Conversely, as \(\alpha\) decreases, the risk of the incorrect rejection of the key points that should properly be used in the template matching increases. Only the sets of the key points that were not excluded by the equation (4) are used for the template matching.
2.3 Detection of unfinished points

Following the judgment of the progress in the work, it is necessary to detect the unfinished points of each task. The subtraction processing can be applied, but the subtraction processing requires input images to be transformed correctly so that the coordinates of the key points of both images may overlap. The following procedures are used to obtain a matrix for executing this transformation.

2.3.1 Determination of an affine matrix

Affine transformation is used for the simultaneous scaling and rotation of images. It can minimize the discrepancy between the key points of the input and template images. The least square method is used to obtain an affine matrix. The coordinates of the key point \( p_k \) \( (k = 1, 2, \ldots, n) \) of an input image are \( (i_k, j_k) \) and the coordinates of the key point \( q_k \) of a template are \( (u_k, v_k) \).

Let represent the two-dimensional affine matrix in equation (5).

\[
A = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix}
\]  

(5)

The relationship of the coincident key points of the input and template images is expressed as:

\[
\begin{bmatrix} u_k \\ v_k \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} i_k \\ j_k \\ 1 \end{bmatrix} = \begin{bmatrix} a_i + b_j + c - u_k \\ d_i + e_j + f - v_k \\ 1 \end{bmatrix}
\]  

(6)

\[
\delta_{ak} = ai_k + bj_k + c - u_k
\]  

(7)

Where, \( \delta_{ak} \) and \( \delta_{ik} \) are error terms that satisfy the equal sign in equation (6). The sum of the squares of the error term \( \delta_{ak} \) is expressed in equation (8).

\[
\Phi_u = \sum_{k=1}^{n} \delta_{ak}^2 = \sum_{k=1}^{n} (ai_k + bj_k + c - u_k)^2
\]  

(8)

The elements \( a, b, c \) of the affine matrix \( A \) are determined so that the error terms \( \Phi_u \) may be minimized. By using the partial differentiation of \( \Phi_u \), the above conditions are expressed in the following equations.

\[
\frac{\partial \Phi_u}{\partial a} = \sum_{k=1}^{n} 2i_k(ai_k + bj_k + c - u_k) = 0
\]  

(9)

\[
\frac{\partial \Phi_u}{\partial b} = \sum_{k=1}^{n} 2j_k(ai_k + bj_k + c - u_k) = 0
\]  

(10)

\[
\frac{\partial \Phi_u}{\partial c} = \sum_{k=1}^{n} 2(ai_k + bj_k + c - u_k) = 0
\]  

(11)

Equations (9), (10), (11) are expressed as:

\[
\begin{bmatrix} \sum_{k=1}^{n} i_k^2 & \sum_{k=1}^{n} i_kj_k & \sum_{k=1}^{n} i_k \\ \sum_{k=1}^{n} i_kj_k & \sum_{k=1}^{n} j_k^2 & \sum_{k=1}^{n} j_k \\ \sum_{k=1}^{n} i_k & \sum_{k=1}^{n} j_k & n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} u_ki_k \\ \sum_{k=1}^{n} u_kj_k \\ \sum_{k=1}^{n} u_k \end{bmatrix}
\]  

(12)

Let the foregoing be replaced with a symbol \( R_u \) and \( S_u \).

\[
R_u = \begin{bmatrix} \sum_{k=1}^{n} i_k^2 & \sum_{k=1}^{n} i_kj_k & \sum_{k=1}^{n} i_k \\ \sum_{k=1}^{n} i_kj_k & \sum_{k=1}^{n} j_k^2 & \sum_{k=1}^{n} j_k \\ \sum_{k=1}^{n} i_k & \sum_{k=1}^{n} j_k & n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} u_ki_k \\ \sum_{k=1}^{n} u_kj_k \\ \sum_{k=1}^{n} u_k \end{bmatrix}
\]  

\[
S_u = \begin{bmatrix} \sum_{k=1}^{n} i_k^2 & \sum_{k=1}^{n} i_kj_k & \sum_{k=1}^{n} i_k \\ \sum_{k=1}^{n} i_kj_k & \sum_{k=1}^{n} j_k^2 & \sum_{k=1}^{n} j_k \\ \sum_{k=1}^{n} i_k & \sum_{k=1}^{n} j_k & n \end{bmatrix} \begin{bmatrix} u_k \\ v_k \\ 1 \end{bmatrix}
\]  

(13)

Then the elements \( (a, b, c) \) are expressed as:

\[
\begin{bmatrix} a \\ b \\ c \end{bmatrix} = R_u^{-1} S_u
\]  

The elements \( (d, e, f) \) are given by replacing \( u_k \) with \( v_k \) in equation (12).

In general, all the key points of the input image are not always correctly associated with those of the template in equation (4). Accordingly the tolerance \( \beta \) is used to eliminate the non-coincident key points associated with the wrong key points of the template. To this end, we correct the error terms in consideration of their absolute values. As an error term increases, the weight of the error term decreases. This correction of the weight is expressed as:

\[
w_k(\delta_{ak}) = \left\{ \begin{array}{ll} \left[ 1 - \left( \frac{\delta_{ak}}{\beta} \right)^2 \right] & (|\delta_{ak}| \leq \beta) \\ 0 & (|\delta_{ak}| > \beta) \end{array} \right.
\]  

(14)

If the error term \( |\delta_{ak}| \) exceeds the tolerance \( \beta \), then the corresponding key points are eliminated from being considered. If the error term \( |\delta_{ak}| \) falls below the tolerance \( \beta \), as the error term \( |\delta_{ak}| \) is decreased, its weight is increased. The weight of the error term \( \delta_{ak} \) is corrected in the same manner and represented by \( w_k(\delta_{ak}) \) in equation (14).

Using the weights of the key points, equation (5) is transformed into:

\[
A' = \begin{bmatrix} a' \\ b' \\ c' \\ 0 \\ 0 \\ 1 \end{bmatrix}
\]  

(15)

\[
\begin{bmatrix} \sum_{k=1}^{n} w_k(\delta_{ak}) i_k^2 & \sum_{k=1}^{n} w_k(\delta_{ak}) i_k j_k & \sum_{k=1}^{n} w_k(\delta_{ak}) i_k \\ \sum_{k=1}^{n} w_k(\delta_{ak}) i_k j_k & \sum_{k=1}^{n} w_k(\delta_{ak}) j_k^2 & \sum_{k=1}^{n} w_k(\delta_{ak}) j_k \\ \sum_{k=1}^{n} w_k(\delta_{ak}) i_k & \sum_{k=1}^{n} w_k(\delta_{ak}) j_k & \sum_{k=1}^{n} w_k(\delta_{ak}) \end{bmatrix} \begin{bmatrix} a' \\ b' \\ c' \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} w_k(\delta_{ak}) u_ki_k \\ \sum_{k=1}^{n} w_k(\delta_{ak}) u_kj_k \\ \sum_{k=1}^{n} w_k(\delta_{ak}) \end{bmatrix}
\]  

(16)

The elements \( (a', b', c') \) are represented by replacing \( u_k \) with \( v_k \) in equation (16).

The affine transformation matrix \( A' \) is obtained by solving the above equations. The matrix \( A' \) expresses the coordinates converted from the unconverted coordinates. However, this procedure disadvantageously nullifies the points after being transformed for the sake of scale conversion of a fractional part. In order to avoid this problem, the inverse transformation is executed. The inverse matrix of \( A' \) is expressed as:
The work environments, such as the brightness of a work place, and the color of facilities. Accordingly, we use the distinction analysis method (method of Otsu [4]) to determine the threshold value automatically. It uses luminance distribution that is calculated from the image after the background subtraction was performed. The threshold value is determined so as to maximize separation metrics. The separation metrics represent a ratio of between-class variance to within-class variance in the luminance distribution.

3. System configuration

This system uses the client-server architecture with the aim of reducing processing time and the cost of development. Fig.4 shows a conceptual diagram of this system. A server is located at a practical skill training center and a client is operated by a worker at a worksite. The server has a high-speed arithmetic unit and stores template images used in the template matching and operation manuals. The client has minimum functions and its main functions are to take pictures using a camera, to display operation manuals and to transmit data to the server.

The client photographs an input image of the front view of a facility or a tool. The input image is automatically sent to the server and high-speed image processing is carried out by the server. The progress in the maintenance work is judged based on the results of image processing.

An operation manual is transmitted to the client. The client need not store such data as template images and operation manuals and the complicated processing such as template matching is performed by the server. Accordingly, high processing power is not required for the client and this contributes to the reduction in the cost of the development of the system. Fig.4 shows the main functions of our client server system.

4. Trial operation

Among all functions of this system for skill training for maintenance, only the functions concerning to the proposed methods are explained. This system was tested to judge the progress in the maintenance work of a typical centrifugal pump many of which are used at chemical plants. The maintenance work concerns the disassembly and overhauling of the pump, which is composed of eight parts, i.e. (1) a casing, (2) an impeller, (3) a ground cover, (4) a ball bearing, (5) a mechanical seal, (6) a bearing cover, (7) a ball bearing and (8) a main shaft. Template images were prepared as the standard of the acceptance level in each template is shown in Table 1. The upper-row of the table shows the number of key points before the exclusion of the non-coincident key points. It is found that a number of non-coincident key points are included. Our system gives precedence to a template with a smaller serial number, but a wrong template was chosen since it clearly has the largest number of coincident key points.

We introduced a method of excluding the non-coincident key points, and this was correctly chosen. The sets of key points with the same characteristic quantities are found by SURF. The number of key points in each template is shown in Table 1. The upper-row of the table shows the number of key points before the exclusion of the non-coincident key points. It is found that a number of non-coincident key points are included. Our system gives precedence to a template with a smaller serial number, but a wrong template, i.e. the template of the 3rd part was chosen. The bottom row of the table shows the number of coincident key points after being excluded in the processing expressed by the equation (4). The 4th template was correctly chosen since it clearly has the largest number of coincident key points.
points. For all the key points, the method checks a gap between the positions of the input and template images. Subsequently the method excludes the key points having the large gap between the locations of the images as non-coincident key points. This trial operation shows that the method allowed us to obtain a clear difference between the coincidence rate of the correct template and that of the other templates. The risk of wrong identification was reduced. However, there is still a problem that the camera angle may vary among individuals. The discrepancy in the camera angle prevents the key points or unfinished points from being detected. This problem will be handled in our following research.

Following to the judgment of the progress in the work, operation manual is shown to the worker. Explanation of the operational manual of this pump is left out since its format is the same to the Fig. 2.

4.2 Detection of unfinished points
As an example, a case of 5th part is shown. The input image of the pump is shown in Fig. 8. The template of the 5th part was collectry chosen by the template matching since it had the largest number of coincident key points. 350 key points were detected in the input image and 307 key points were detected in the 5th template image. Among those key points, there were 23 coincident key points. An affine matrix was given by:

$$A' = \begin{pmatrix} 0.9926 & -0.0214 & 20.4055 \\ 0.0214 & 0.9926 & -19.4475 \\ 0 & 0 & 1 \end{pmatrix} \quad (19)$$

Its inverse matrix is given by:

$$A'^{-1} = \begin{pmatrix} 1.0070 & 0.0217 & -20.1264 \\ -0.0217 & 1.0070 & 20.0264 \\ 0 & 0 & 1 \end{pmatrix} \quad (20)$$

The input image in Fig. 8 was transformed by the inverse matrix $A'^{-1}$. Subsequently the subtraction processing was performed. Fig. 10 shows an image after the subtraction. A threshold value for the binarization was determined by the distinction analysis method and was set at 71.0. Fig. 11 shows the result of the binarization and it was found that this system correctly detected an unfinished point (mechanical seal) in the 5th part. The result shown in Fig. 11 is shown to the worker at the work site.

4.3 Processing time
The input images are sent through the Internet to the server. The performance of this system using SURF was compared with the prototype system using SSD. The number of templates was increased to 29. Then, the processing time required for template matching was measured. The processing time of the SSD system increased steeply with an increase in the number of template images. Its processing time for the template matching
of the 29 templates was 130 seconds. The SURF system con-
trolled the processing time better. It took 18 seconds to com-
plete the template matching of the 29 templates. The processing
time was shortened by 85%. The difference in the processing
time between the two systems was remarkable in the case where
the number of templates was more than 20.

5. Conclusion
This research developed the system to judge work progress
from the pictures taken of the facilities to be maintained. The
main results of this research are shown below.

1. A method was proposed to remove non-coincident key
points to improve the accuracy of template matching.

2. The subtraction processing was added to automatically de-
tect unfinished points.

3. This system uses the client-server architecture, which
achieves a realizable system in terms of the time and cost
of image processing, work procedures and centralized data
management.

All of the above features were verified by the experiment per-
formed on the prototype system using SSD and the new system
using SURF.

References
[1] Uchida M., 1987. Current topics of damage and repair of chem-
ical plant equipment, Journal of the Japan Welding Society,
56(6), 358-367.
[2] Izumi K., Yasuhiro Kajihara, et al., 2013. A support system for
the tube replacement operation of a heat exchanger, Journal of
the Society of Plant Engineers Japan, 25(2), 46-51.
[3] Bay H., et al., 2006. SURF: Speed-Up Robust Features, 9th
Euro. Conf. on Computer Vision, 404-417.
[4] Otsu N., 1979. A threshold selection method from gray-level
histogram, IEEE Trans. Systems, Man and Cybernetics, SMC-
9, 62-69.

Ying Yu
is now a PhD candidate of the Graduate School of
System Design, Tokyo Metropolitan University. She re-
ceived her master degree in 2014 from Okayama Shoka
University, Japan. She is studying in the field of indus-
trial engineering and industrial technology.

Yasuhiko Kajihara (Member)
is a professor in the graduate school of system design,
Tokyo Metropolitan University. He received a PhD de-
gree in 1990 from Okayama University and received a
master's degree from Tokyo Institute of Technology in
1985. His current research interests include industrial en-
gineering and industrial technology.

Akashi Tomoyuki
joined Sankyu Inc. in 1980. His specialized field is
manufacturing pressure containers in the petrochemical
plant facilities. He is now a director of the maintenance
center at Chiba branch.

Izumi Kazuhiko
joined Sankyu Inc. in 1981. His specialized fields are
industrial engineering and information system. He re-
ceived a PhD degree in 2013 from Tokyo Metropolitan
University and received a master’s degree in 2007 from
Waseda University. He is now on loan to infoSense Corp.
and a director of the infrastructure service division.

Jindai Mitsuharu
is a professor in the Graduate School of Science and
Engineering, University of Toyama. He received a PhD
degree and a master’s degree from Okayama University
in 1999 and 1996, respectively. His current research in-
terests include human-robot interaction and image pro-
cessing.