Adaptive Annotation Using a Human-Robot Interface System PARTNER

Masaya Yamashita　　Shigeyuki Sakane

Department of Industrial and Systems Engineering
Chuo University
1-13-27 Kasuga, Bunkyo-ku, Tokyo, 112-8551 JAPAN
E-mail: {yamasita, sakane}@indsys.chuo-u.ac.jp

Abstract

We are developing a human-robot interface system, PARTNER, that takes into account the flexibility of the Augmented Reality approach. The prototype consists of a projector subsystem for information display and a real-time tracking vision subsystem to recognize the environment and the human's action. For teaching robot tasks, three kinds of interaction have been developed: Virtual Operational Panel, Interactive Image Panel, and Interactive Hand Pointer. This paper presents the fourth function, Adaptive Annotation Function (AAF) for teaching and assisting the human's task. Based on a state transition diagram, the system generates annotations adaptively to the situation by monitoring not only changes of the object's geometry but also the operator's action required for achieving the task. To detect changes of the environment, we have implemented structured light rangefinder capabilities using components of the prototype system. Experimental results of an unfolding task of a portable OHP device demonstrate the usefulness of the proposed system.

1 Introduction

To extend the application field for next-generation robot systems, interfaces that are easier to use and more friendly need to be developed for various levels of human-robot interaction.

In the field of human-computer interfaces, much attention has recently been paid to Augmented Reality [1][2] and Mixed Reality [3] systems which can enhance a human's daily life by blending multi-modal information with the real world. The approach suggests a promising direction for the development of the human-robot interface. By taking into account the flexibility of the AR approach, especially the attempt of DigitalDesk [4], we have developed a human-robot interface system, PARTNER (Projector-based AR for Teaching NEw tasks between human and Robot). The prototype consists of a projector subsystem for information display and a real-time tracking vision subsystem for recognizing the task environment and the human operator's actions.

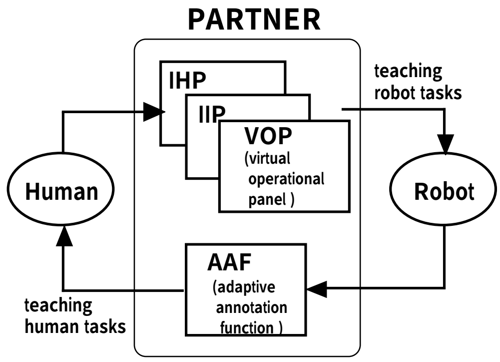

Figure 1: Adaptive Annotation Function and other modules in PARTNER

In PARTNER, three interactive functions have been developed: Virtual Operational Panel (VOP), Interactive Image Panel (IIP), and Interactive Hand Pointer (IHP) [5],[6]. These functions are basically used for teaching robot tasks. In contrast, this paper presents the fourth function, Adaptive Annotation Function (AAF) as shown in Figure 1. We can use AAF for teaching and assisting the human operator's task via projecting annotations adaptively to the situations which may occur during the tasks.

2 Previous Studies

Although some AR systems have been developed for assisting human tasks in manufacturing and repair, most of the systems use see-through head mounted displays (HMD) or monitor displays.

Feiner et al. developed KARMA, a typical AR system which helps an end user to maintain a laser printer [7]. The system uses a see-through HMD and allows the user to see the 3D geometric model (wire-frame model) of the printer and to locate/identify the toner cartridge and paper tray which need to be handled. Since the expected movement and the results of the parts are superimposed on the real printer, it is easy for the user to understand the necessary operations. However, since the system uses HMDs and the position of the user's head may change relative to the printer, the head position must be tracked in real-time to correctly superimpose the 3D model on the real object.

Caudell and Mizell developed an AR system for guiding a technician in building a wiring harness used for an airplane's electrical systems [8]. Also in this system, the user must wear a see-through HMD to see the layout diagrams for the wire bundles superimposed on the layout board.

Ban et al. developed an AR system for supporting inspection work of electronic parts. The system superimposed visual markers, visual commands, and virtual instruments on an electronic board. They used a monitor display and a half mirror, instead of an HMD, to superimpose a screen image of the display on the actual electronic boards. Therefore the user's head position is constrained to the display and the mirror.

DigitalDesk [4] is a well-known example of the projector-based AR system. However, the system supports a human's desk work by projecting a virtual calculator, virtual drawings, etc. and monitoring the user's hand gestures. Sato et al. [10] developed a system for supporting a human's fabrication task such as assembly of electronic circuits. They also used a projector-based AR and monitored the user's gestures for effective support.

3 Adaptive Annotation Function of PARTNER

In contrast to the above AR-based annotation systems, our proposed system (PARTNER/AAF) has the following features:

1. Since PARTNER uses a projector, the user does not need to wear headgear such as see-through HMDs or to watch a monitor display. The system allows multiple users to share annotations.

2. Both monitoring of geometrical changes of the task environment and of the user's actions are utilized for adaptive annotations. We use a state transition diagram for adaptively generating annotations as shown in Figure 2.

3. The system utilizes rangefinder capabilities for detection of geometrical changes of the environment. Since the projection subsystem and the visual tracking are components of PARTNER, we use them for implementing structured light rangefinders such as random-dot stereo and light-striped rangefinder without additional hardware.

4. Since the user may act differently depending of the knowledge of the task, the annotation system should adapt to the user by monitoring his/her actions in the task. As an example, the system generates an additional annotation for a novice user who cannot puzzle out how to operate. The system provides flexible guidance.

5. The system uses infrared images for reliably detecting and monitoring the user's hand actions. The infrared images are robust against background clutter, change of color, and texture.

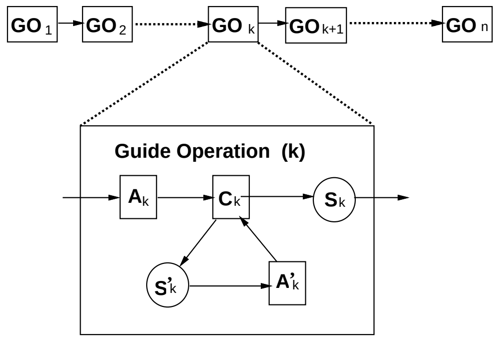

Figure 2: State transition diagram of guiding operations and adaptive annotations (GO_k:guide operation, A_k:annotation, C_k:check of status, S_k:status)

In the following experiments, we will show an unfolding task of a portable OHP (over head projector) device. As shown in the processes, PARTNER/AAF provides a virtual and interactive guide book using the real object in the real environment.

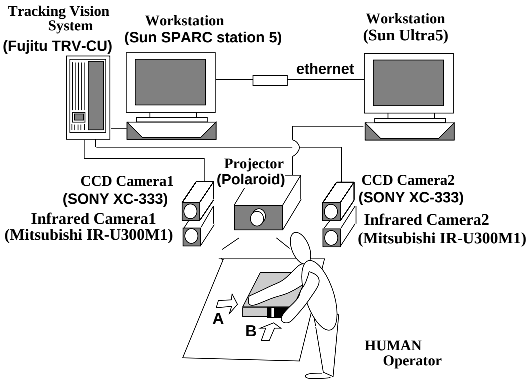

Figure 3: The prototype system

Figure 4: A scene of 3D measurement using the random-dot stereo

4 The Prototype System

4.1 Hardware and Software

Figure 3 shows the hardware of our prototype system. A color tracking vision system (TRV-CU, Fujitsu) [11] tracks template images in real-time based on a block matching algorithm. It allows us to track about 500 template images with a size of 8×8 $(pixels)$ in a TV frame, 33 $(msec)$. Live images of the task environment are captured through CCD cameras (G20 and XC-333, SONY). These cameras are used for the structured light rangefinders to localize objects and detect geometric changes of the task environment. A pair of infrared cameras (IR-U300M1, Mitsubishi) are used for monitoring the user's hand action.

We use an LCD (liquid crystal device) projector (Colorview Light, Polaroid) to provide information for the user by projecting virtual panels for the VOP and IIP [5], a mark for the IHP[6], and annotations for the AAF. The resolution of the projector is $800 \times 600(pixels)$ and the light output is 500 ANSI lumen. The projector is connected to a workstation (Ultra5, Sun), an X-window server for the projection.

For the software in our prototype system we used EusLisp [12], an object-oriented lisp language developed for robotic applications at the Electrotechnical Laboratory. EusLisp allows the user to call X-window library functions and to integrate the library functions of the devices as lisp functions.

4.2 Structured light rangefinders in PARTNER

For the adaptive annotation function, we have to detect and monitor changes of the task environment and the user's action. Therefore, we have implemented the random-dot stereo and light-striped rangefinder using the projector and the vision system, which are current components of PARTNER. Typical uses of the rangefinders are the following:

1. Extracts rough 3D information of the task environment using the random-dot stereo and determines the principal axis of the blob region.
2. Measures 3D data more precisely using the light-striped rangefinder where the light-stripes are perpendicular to the principal axes.

This process is used for fitting the 3D model of the target object to the environment.

4.2.1 Random-dot stereo

A problem in stereo vision is to find correspondences accurately in a pair of images. By projecting random-dots onto the scene we can improve the uniqueness of the image and decrease incorrect matchings in the stereo search. In our system, a template subimage with a square shape is defined in the right image and the template is searched in the left image by correlation. The size of the input image is 640×480 $(pixels)$, and the size of the template is 32×32 $(pixels)$ at one-fourth resolution. The number of search areas used in the left image is 69×49 $(blocks)$. The array elements with high correlation values are merged into regions and labels are given to the regions. Figure 4 shows an example scene of the random-dot stereo.

4.2.2 Light-striped rangefinder

We have also implemented a light-striped rangefinder using the projector and the tracking vision system. To

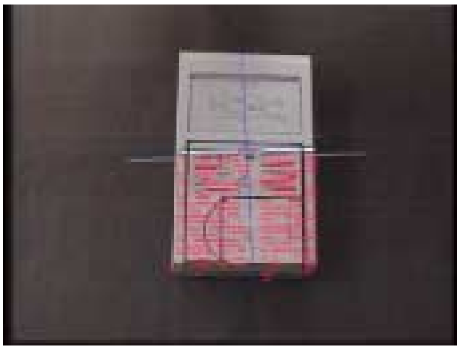

Figure 5: A scene of 3D measurement using the light-striped rangefinder (The principal axis is superimposed on the image.)

Figure 6: Unfolding of a portable OHP device is a difficult task for a new user without consulting the manual.

determine the position of the striped light, we used four templates of the light which correspond to the directions of 0, 45, 90, and 135 degrees. Since position, orientation, length, and width of a striped light can be controlled by software, we can adaptively project the light onto the scene. For example, we can limit the region of interest around the target objects using the results of the random-dot stereo.

Before starting the measurement, both the projector and the cameras are calibrated for stereo measurement with respect to the world coordinates. For geometric modeling of the projector, we used a perspective transform specified by seven parameters of position, orientation of the optical center, and a scale factor.

5 Experiments of guiding an unfolding task of a portable OHP device

Unfolding of a portable OHP device is often a difficult task for a new user without consulting the user's manual (Fig. 6). We conducted experiments of this task to evaluate the proposed system.

We use the state transition diagram (Fig.2) of guiding operations including adaptive annotations for the unfolding task. Each guiding operation GO_k begins with the initial annotation A_k and starts checking (or monitoring) the task status. If the desired status is achieved, then the task state transits to S_k. Otherwise, the second annotation is generated to prompt the desired operation or to warn the user of the wrong operation.

The unfolding task requires the following steps:

(1) Unlock the device by moving the head-lock lever,
(2) Lift up the main body,
(3) Slide up the head part and rotate it to be horizontal,
(4) Pull the mirror, lift it up, and rotate it to a suitable direction,
(5) Open the power cord box, pull the cord out, and connect the plug to a power supply,
(6) Then turn the switch on.

The guiding operation GO_1 includes the following steps:

1. Acquires 3D range data using the random-dot stereo and the light-striped rangefinders
2. Locates the geometric model of the device in the real environment
3. Locates the positions of lock levers
4. Generates annotations (A_1) for the unlock operation by indicating the position by an arrow.
5. Detects the user's hand in the infrared image
6. When the position is not close to the unlock lever in a certain time (C_1), the system generates an additional annotation (A_1') to prompt the operator's unlock action.

In processing the infrared images, the system has two modes: (1) thinning of the hand region to extract fingers, and (2) tracking of the fingertip in real-time for monitoring the hand behavior. As shown in Fig.8, the thinning process produces a skeleton of the fingers so that the hand region may be identified reliably. In

Figure 7: An example of annotation to guide the unfolding task

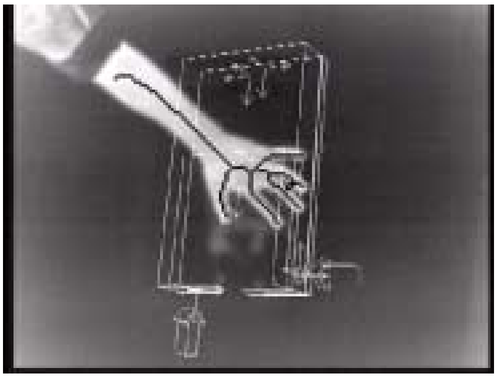

Figure 8: An infrared image for detecting the human hand. The thinning results and the object model are superimposed.

the process C_1 to check whether the main body is properly opened, the distance between the hand and the unlock lever is monitored within a certain time interval. If the distance is not close enough, the user may not have identified the unlock lever yet, so the system generates additional annotations to prompt the operation (Fig.9).

In the guiding operation GO_2 which corresponds to the initial unfolding of the main body, the following steps are provided:

1. The annotation message, "Unfold the main body" (A_2) is projected.

2. The height of the edge of the main body is measured (C_2) using the light-striped rangefinder.

3. If the height of the main body is not sufficient compared with the geometric model in the expected state, the second annotation, "Lift up more" (A_2') is projected.

4. When the state (the body angle is sufficient) (S_2) is satisfied, the GO_3 ends and the system transits to GO_4.

Figure 9: Adaptive annotation by monitoring the user's action

Figure 10: The message, "Unfold the main body" is projected

Figure 12 shows the scene of checking the height of the main body using the light-striped rangefinder. Since the user's hand may be around the body, the system projects a couple of striped lights and calculates the 3D data (profile) along the lights. Then, from the candidate profiles, we select one which has no overlap with the hand region which can be estimated from the infrared image of the scene. Figure 13 shows two cases of the profile data when the lifting operation is incorrect and correct, respectively. Since

Figure 11: The message, "Lift up more" is projected when the height of the main body is smaller than the model.

Figure 12: A striped light is projected to check the height of the main body.

the system has the geometric model of the target object, we can predict the correct profile as shown in the figure and compare it to the measured data to check the operation is correctly performed.

The guiding operation GO_3 starts with the annotation message, "Pull up the head part". This operation is slightly difficult for a novice user since the head part must be slid up along the slant angle and it must be rotated to be horizontal.

1. The annotation message, "Slide up the head part" (A_3) is projected.
2. The attitude of the head part is verified (C_3) using the light-striped rangefinder (Figure 15). The profile data of the head part is shown in Figure 16. Figure 16(a) shows an incorrect status and Fig.16(b) shows a correct status, respectively.
3. If the head part is not properly unfolded, the second annotation message, "Rotate the head part to be horizontal" is projected (A_3').

Although we omit here a description of the subsequent steps, the process is continued in a similar way.

6 Conclusions

We presented the Adaptive Annotation Function in the human-robot interface system PARTNER. We verified the basic functions through the task of unfolding a portable OHP device. In contrast to the previous AR-based annotation systems, the new function has the following features:

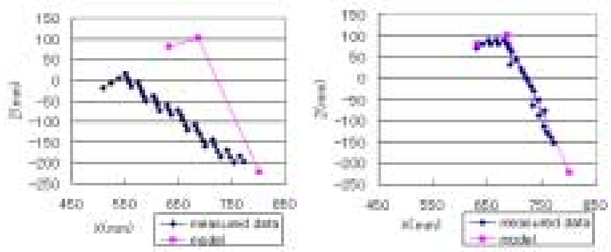

Figure 13: Comparison of the profile of the main body with the model: (a) incorrect operation, (b) correct operation

- Since the PARTNER uses a projector, the operator need not ware HMDs and multiple users are allowed to share the annotation.
- Adaptive annotation has been implemented by monitoring both the change of the object geometry and the operator's action.
- The function utilizes the structured light rangefinders which are implemented using components of PARTNER.
- We used infrared images for reliable detection and monitoring of the operator's action.

In recent years, much progress has been made in projector technology such as DLPs, and projectors with a compact size, high intensity, and high resolution are commercially available today. In the near future, such devices will be used for implementing projector-based annotation systems.

Figure 14: The message, "Rotate the head part of be horizontal" is projected.

Figure 15: The state that the head part is properly rotated to be horizontal.

Our plans for future research include (1) improvement of the method for checking the operator's action, (2) integration of more flexible interactions such as handing over tools between human and robot, and (3) integration of voice-based annotations.

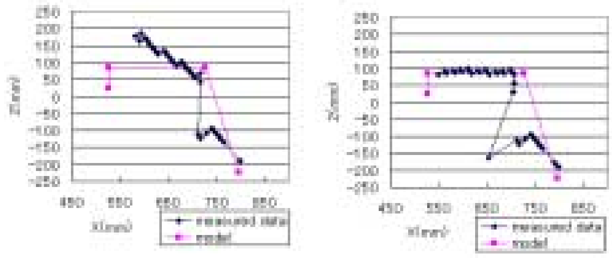

Figure 16: Comparison of the profile of the head part with the model : (a) incorrect operation, (b) correct operation

References

[1] Computer-augmented environments: Back to the real world, *Communication of ACM*,pp.24-97,1993

[2] R.T.Azuma: "A survey of augmented reality," *PRESENCE*, Vol.6, No.4, pp.355-385,1997.

[3] Y.Ohta, H.Tamura: *Mixed Reality -Merging real and virtual worlds*, Springer, 1999.

[4] P.D.Wellner: "Interacting with paper on the DigitalDesk," *Communications of the ACM,* Vol.36, No.7, pp.86-97, 1993.

[5] M.Terashima, S.Sakane: "A human-robot interface using an extended Digital Desk" *Proc. IEEE Int. Conf. Robotics and Automation,* pp. 2874-2880, 1999.

[6] S.Sato, S.Sakane "A human-robot interface using an interactive hand pointer that projects a mark in the real work space," *Proc. 2000 IEEE Int. Conf. Robotics and Automation,* pp.589-595, 2000.

[7] S.Feiner, B.MacIntyre, D. Seligmann,"Knowledge-based augmented reality," *Communications of the ACM*, Vol.36, No.7, pp.53-62, 1993.

[8] T.P.Caudell, D.W.Mizell, "Augmented reality: An application of head-up display technology to manual manufacturing processes," *Proc. Hawaii Int. Conf. System Science*, pp.659-669, 1992.

[9] Y.Ban, S.Wada, K.Sato, and K.Chihara,"A desktop environment for electronic inspection by augmented reality," *Proc. Japan-USA Symposium on Flexible Automation*, pp.1277-1280, 1998.

[10] T.Sato, E.Inoue, and H.Mizoguchi, "Support in situ for operation of fabrication," *Proc. ROBOMEC'95*, pp.535-538, 1995 (in Japanese).

[11] T.Morita, N.Sawasaki, T.Uchiyama, and M.Sato: "Color tracking vision system" *Proc. of the 14th Annual Conference of the RSJ,* pp. 279-280, 1996 (in Japanese).

[12] T.Matsui: "Multithread concurrent object-oriented language EusLisp with geometric modeling facilities" *Journal of the RSJ*, Vol. 14, No.5, pp. 650-654, 1996 (in Japanese).