

Task Support System by Displaying Instructional Video onto AR Workspace

Michihiko Goto*
Keio University

Yuko Uematsu†
Keio University

Hideo Saito‡
Keio University

Shuji Senda§
NEC Corporation

Akihiko Iketani¶
NEC Corporation

ABSTRACT

This paper presents an instructional support system based on augmented reality (AR). This system helps a user to work intuitively by overlaying visual information in the same way of a navigation system. In usual AR systems, the contents to be overlaid onto real space are created with 3D Computer Graphics. In most cases, such contents are newly created according to applications. However, there are many 2D videos that show how to take apart or build electric appliances and PCs, how to cook, etc. Therefore, our system employs such existing 2D videos as instructional videos. By transforming an instructional video to display, according to the user's view, and by overlaying the video onto the user's view space, the proposed system intuitively provides the user with visual guidance. In order to avoid the problem that the display of the instructional video and the user's view may be visually confused, we add various visual effects to the instructional video, such as transparency and enhancement of contours. By dividing the instructional video into sections according to the operations to be carried out in order to complete a certain task, we ensure that the user can interactively move to the next step in the instructional video after a certain operation is completed. Therefore, the user can carry on with the task at his/her own pace. In the usability test, users evaluated the use of the instructional video in our system through two tasks: a task involving building blocks and an origami task. As a result, we found that a user's visibility improves when the instructional video is transformed to display according to his/her view. Further, for the evaluation of visual effects, we can classify these effects according to the task and obtain the guideline for the use of our system as an instructional support system for performing various other tasks.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Training, help, and documentation

1 INTRODUCTION

There is a field of research that concentrates on the use of augmented reality (AR) for providing instructional support for tasks that need to be performed manually by a human being. We can simultaneously view multiple virtual objects in the real world by using an AR-based system. Therefore, it is possible to present the information required for performing manual tasks intuitively; this information is applied to various manual tasks such as navigation, assembling a puzzle, and playing various instruments. In the case of navigation, the system displays the information related to a route by using camera-captured images [10]. In the case of assembling a

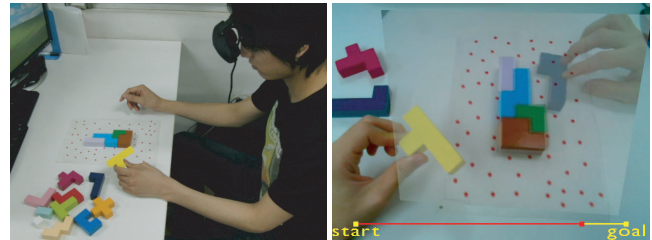


Figure 1: (Left) Work environment and (Right) image view through the HMD

puzzle, the system overlays a necessary part directly onto the relevant part of the entire image and displays an annotation in the appropriate place. By using this system, the user can assemble a puzzle intuitively and smoothly. Further, in order to illustrate the case of playing various instruments, we provide an example of the AR-based instructional support system developed by Motokawa et al.; this system provides instructions on how to play the guitar [9]. This system overlays a virtual hand model onto the position that should hold the strings. Even a beginner can learn to play the guitar easily by using this system.

An AR-based system has three functions: (1) the creation of some virtual objects as additional contents, (2) the estimation of the geometric position between the camera and the environment, and (3) the adjustment of virtual objects positions in the real world by using the estimated geometric position. By carrying out the estimation of the geometric position and adjusting virtual objects positions in real time, an AR-based system can display additional virtual objects in the real world without making the simulated environment look strange.

Thus far, there have been a number of researches related the provision of intuitive and interactive instructional support for various tasks by using AR concepts. Reiners et al. [12] described an AR demonstrator for the task of door lock assembly into a car door. Flagg et al. [4] proposed an interactive system as a painting guide; this system projects the task of painting by dividing the canvas into several sections. Salonen et al. [14] suggested the use of an AR-based system for supporting assembly tasks. Nilsson et al. [11] developed an AR-based interactive system and confirmed the usability of this system in a hospital (the hospital was considered to be an example of a limited workspace). Steven et al. [5] applied a large-scale AR-based instructional support system to a mechanical task performed by the military. Robertson et al. [13] evaluated the effects of a graphical context in a Lego block placement task when the graphics were located outside of the task area. As a result of the evaluation, they concluded that registered AR outperformed non-registered AR in the case of tasks in which the placement position is important, such as the Lego block placement task. A.Tang et al. [15] evaluated the effect of the error rate for assembly tasks and measured mental effort by using AR instruction. As a result, they indicated that the assistant by using AR instruction is better than the other assistant by using printed manual or computer assisted instruction.

*e-mail: goto@hvrl.ics.keio.ac.jp

†e-mail: yu-ko@hvrl.ics.keio.ac.jp

‡e-mail: saito@hvrl.ics.keio.ac.jp

§e-mail: s-senda@ap.jp.nec.com

¶e-mail: iketani@cp.jp.nec.com

Most of these researches use computer graphics, images, signage, and so on as additional AR contents. However, new contents have to be created for each situation. In this paper, therefore, we propose a system that does not create new contents but directly uses existing model instructional videos as the additional contents. The prototype design of AR task support system by using instructional video is shown in Fig. 1.

The following is an example of the use of an instructional video to perform a task: When workers learn how to assemble a part in factories, they play an instructional video, and imitate the shown guideline. Such videos are created by companies or provided on the Internet [1, 2]. However, in order to perform the task while watching the video, they need to alternate between looking at their hands and watching the video. In other words, there is constant eye movement during the process. In CSCW field, these are a number of researches for improving the user's understanding condition by projecting the other's workspace onto the user's workspace. (e.g. VIDEODRAW of J.Tang et al. [16], TeamWorkStation of Ishii [6], Sheared View System of Kuzuoka [8] and so on). If the speed of the video is faster than the workers' speed, they may fall behind on the task or miss some important part of the video while they are looking at their hands and trying to imitate the actions shown in the video. Tversky et al. [17] reported that a simple display of an animation or a movie does not support the user's work very well, because it is often too complex or too fast for the user to understand details of the task in the animation or the movie. Therefore, they insisted that the interactivity is important. If the user can control the speed of the animation and can view and review, stop and start, zoom in and out, and change orientation, it is easier to understand for the animation or movie.

To overcome these issues, our system simultaneously provides the instruction video and the user's hand by overlaying the video onto the user's camera. In addition, in usual case, the user changes the viewpoint of the video in the brain to adjust his/her viewpoint since the viewpoint of the instructional video might be different from the user's viewpoint. However, it may cause confusion. Therefore this function is very useful for the user's intuitive understanding.

2 PROPOSED METHOD

2.1 System Overview

In the proposed system, we use instructional videos that are transformed to display according to the user's view and are overlaid onto the user's view. We suppose that the user has a camera mounted on his/her head. The display style of augmented view is selectable from a stationary display, HMD, mobile phone, and so on. There are two important concepts related to our system.

First, the system adjusts the pace of the instructional video according to the user's pace such that the user can perform the task at his/her own pace. To begin with, a provider needs to divide the instructional video into several parts per step in advance, as shown in Fig. 2. The system repeats the action of a particular step in the video until the user gives a sign to proceed to the next step. Further, the system pauses the video at the end of each step to show the final appearance of the objects used in the step. Therefore, the user can check whether he/she has completed all the actions related to the step. When the user considers the step successful, he/she gives the sign to proceed to the next step. It is also possible to go back to the previous step. Therefore, the user can watch the same step in the video as many times as he/she wishes to and does not have to worry about being left behind by the instructional video, as shown in Fig. 3. (Here, we define an AR-button as one of the user's sign. If the user puts his/her hand in the area of the AR-button, the AR-button reacts by recognizing skin color of human's hand. The example is shown in Fig. 4.)

Second, the system provides the user with an AR view in which

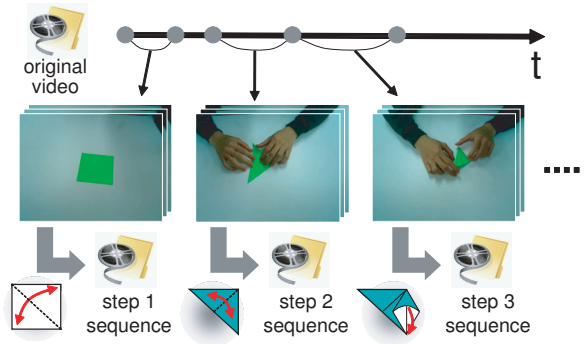


Figure 2: Segmentation of the instructional video

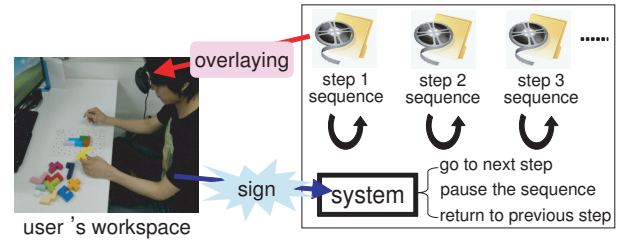


Figure 3: Step-by-step display process of the instructional video in the proposed system

the instructional video is transformed to display according to the user's view and overlaid onto the object that the user is working upon. By transforming the instructional video to display according to the user's view, the system ensures that the user can intuitively understand the operation as if an instructor were teaching beside him/her. Two problems can be identified this concept: (1) the problem of obtaining the geometrical relationship between the instruction video and the user's view and (2) the problem of visual confusion between the instructional video and the user's view.

In order to obtain the geometrical relationship between the instructional video and the user's view, we need to compute two geometrical relationships: (1) the relationship between the view and the object that the user is manipulating and (2) the relationship between the view and the same object in the instructional video. The cause of the visual confusion mentioned above is that the real object the same object in the video are displayed at the same position and at the same time in the AR view, as shown in Fig. 6. Overlaying the instructional video can provide to the user an intuitive understanding of the performed action. However, the visual confusion may prevent a successful imitation of an action. Therefore, we have to maintain the advantage of providing intuitive support and focus on a method in order to resolve the problem of visual confusion.

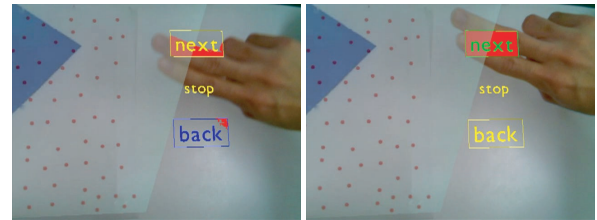


Figure 4: Detection of the AR-button: (Left) is not enough, and (Right) is enough

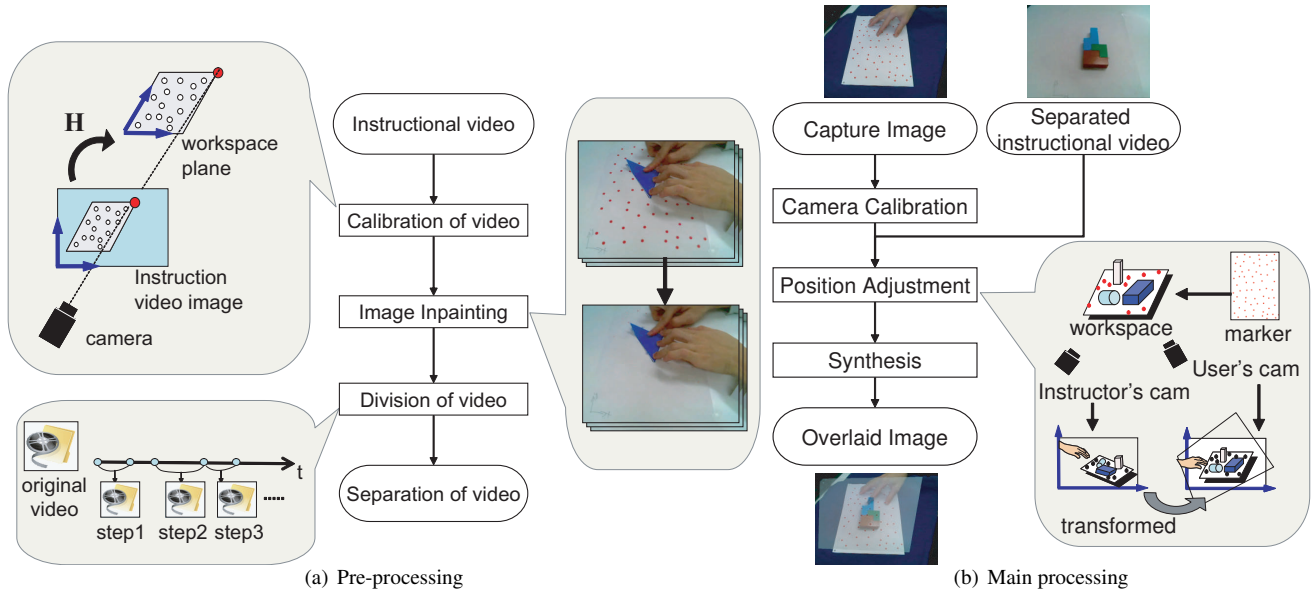


Figure 5: Flow chart of the proposed method

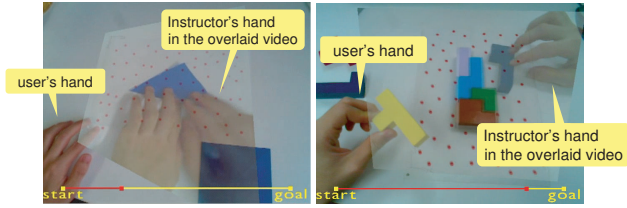


Figure 6: Visual confusion between the instructional video and the user's view: (Left) is the case of making origami, and (Right) is the case of placement block

2.2 Workflow of the proposed method

The flowchart of the proposed method is shown in Fig. 5.

- Preparation of the instructional video (pre-processing)
- Making the imitation of the actions shown in the video (main processing)

In the pre-processing step, we compute the geometrical relationship between an instructor's view and the object in the instructional video, and edit the video accordingly. This step does not need to be performed in real time.

During the main processing, we compute the geometrical relationship between the user's view and the object manipulated by the user. Then, the instructional video is transformed to fit the user's view and applied over the object that the user is working upon. This processing is carried out in real time.

3 PRE-PROCESSING

3.1 Calibration of the instructional video

By computing a homography matrix that represents the geometrical relationship between the user's view (camera) and the real object, the instructional video is transformed. ARToolKit [7] is widely known as the library that can recognize the geometrical relationship between the camera and a target scene by using a rectangular marker. However, it failed when the marker moves out of the image (for example, when the user's view moves frequently) or when

the marker is occluded (e.g., hands and the objects to be worked upon). Therefore, we use the camera tracking method proposed by Uchiyama et al. [18], which can track a part of pattern of a group of dots by identifying each dot based on local arrangement of its neighboring dots. Henceforth, we call this marker a point marker. The point marker is robust against the partial occlusion problem because the marker tracker recognizes the localized positions of the point groups and tracks these points.

We assume that this point marker is placed on the table when the provider captures the instructional video. Because of this assumption, we cannot use an existing videos; however, this is sufficiently acceptable since the provider only needs to print the point marker on a paper and place it in the work environment. In the proposed system, we compute the homography matrix by using that point marker. The geometrical relationship between the view and the object is shown in Fig. 5.

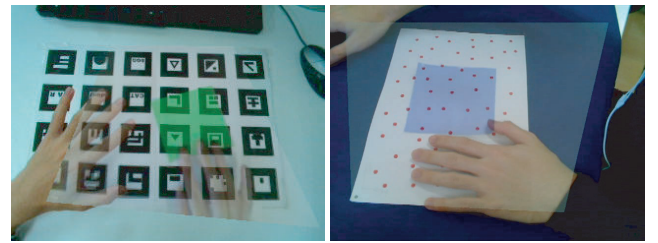


Figure 7: Fitting the marker and the system: (Left) in the case of using the ARToolKit marker and (Right) in the case of using a point marker

3.2 Image inpainting of the points of the point marker

As described before, the proposed system uses a point marker placed on the table for computing the homography matrix. However, users may consider the point marker as an eyesore when he/she watches the overlaid view. Therefore, we need to diminish the appearance of the points.

Image inpainting is an algorithm based on pixel interpolation that considers the continuity of pixel intensity. Therefore, we em-

ploy this algorithm and interpolate the points of the marker to make them invisible in the video, as shown in Fig. 8. In this way, the problem of the point marker being an eyesore is resolved. Since it is difficult to perform the image inpainting in real time, we remove the point marker only in the instructional video during the pre-processing stage.

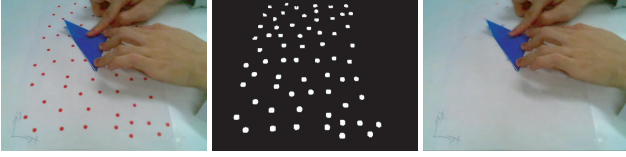


Figure 8: Inpainted image of the instructional video: (Left) input image, (Middle) mask image of the points, and (Right) inpainted image

3.3 Division of the instructional video

The proposed system plays the instructional video according to the required step of operation. This is our first concept the system adjusts the pace of the instructional video according to the user's pace so that the user can carry on the work at his/her own pace. In our current system, the provider manually divides the instruction video into steps by watching the video.

4 MAIN PROCESSING

4.1 Camera calibration

We have to compute a homography matrix as the geometrical relationship between the user's view and the object. The user places the point marker on the table in advance because the proposed system assumes that the user works in the same environment as the one of the instructional video. The homography matrix estimated in the user's view is used in a synthetic step along with the homography matrix estimated in the instructional video's view.

4.2 Position adjustment

In order to fit the video's view to the user's view captured from the camera, we have to transform the image of the instructional video and overlay the video image onto the user's vision. It is necessary to perform this processing in real time while the user works with this system. First, we set up the geometrical relationship between the coordinates of the instructional video and the coordinates of the user's camera.

We define the arbitrary world coordinate of the marker as $\mathbf{x}_1(x_1, y_1)$ and the same position on the video as $\mathbf{x}_2(x_2, y_2)$. In the same way, this position seen from the user's viewpoint is defined as $\mathbf{x}_3(x_3, y_3)$, as shown in Fig. 9. The relations between these coordinates are represented by the following equation:

$$\mathbf{x}_1 \cong \mathbf{H}_{12}\mathbf{x}_2, \quad \mathbf{x}_1 \cong \mathbf{H}_{13}\mathbf{x}_3 \quad (1)$$

\mathbf{H}_{12} is a 3x3 homography matrix between the coordinates of the marker and the coordinates of the instructional video obtained during the pre-processing; \mathbf{H}_{12} does not change as long as the video is captured by a fixed camera. \mathbf{H}_{13} is a 3x3 homography matrix between the coordinates of the marker and the coordinates of the user's camera obtained by main processing; it is necessary to compute this matrix for every frame because the user's camera moves constantly. In order to transform the instructional video to display according to the user's view, we need to compute a new homography \mathbf{H}_{23} between the coordinates of the instructional video and the coordinates of the user's camera, using the following equation:

$$\mathbf{x}_3 \cong \mathbf{H}_{13}^{-1}\mathbf{H}_{12}\mathbf{x}_2 \cong \mathbf{H}_{23}\mathbf{x}_2 \quad (\mathbf{H}_{23} = \mathbf{H}_{13}^{-1}\mathbf{H}_{12}) \quad (2)$$

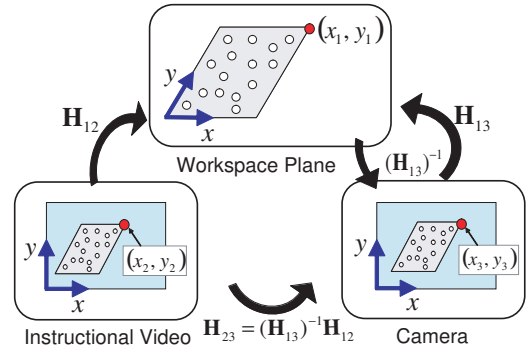


Figure 9: Relationships among the coordinate systems

4.3 Synthetic step

We have to transform the image of the instructional video and overlay the video image onto the user's vision by using the homography matrix \mathbf{H}_{23} computed in the position adjustment step, as shown in Fig. 9.

However, the object manipulated by the user and the object of the video are displayed at the same position and at the same time in the AR view; therefore, there is visual confusion. Hence, we add various visual effects to the instructional video, such as transparency and enhancement of contours. (We will explain these effects in the next section.)

We carry out alpha blending, which varies the luminance values of the instructional video image and the user's camera image. We define the luminance value of an arbitrary coordinate of the transformed instruction video image as $\text{value}_i(R_1, G_1, B_1)$, and the luminance value of the coordinate of the user's camera image as $\text{value}_u(R_2, G_2, B_2)$. The luminance value $\text{value}_s(R, G, B)$ of the coordinate of a synthetic image is as follows:

$$\text{value}_s = \alpha \cdot \text{value}_i + (1.0 - \alpha) \cdot \text{value}_u \quad (0.0 \leq \alpha \leq 1.0) \quad (3)$$

We show the result of the synthetic image in Fig. 10.

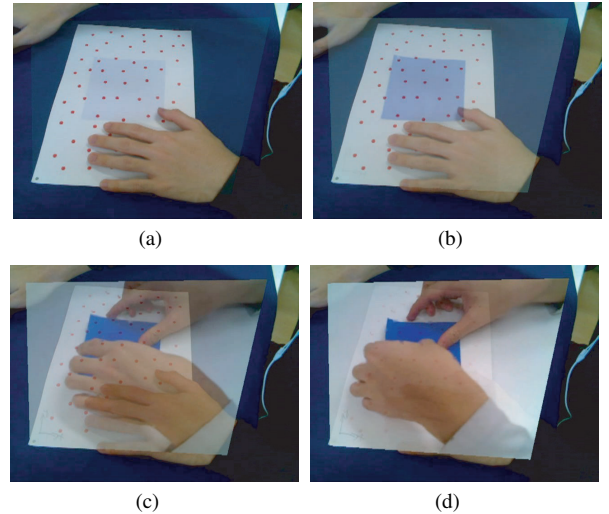


Figure 10: Synthetic images for each transparency value: (a) $\alpha = 0.1$ (prior the user's view), (b) $\alpha = 0.3$ (prior the user's view), (c) $\alpha = 0.7$ (prior the instructional video), and (d) $\alpha = 0.9$ (prior the instructional video)

5 VARIOUS VISUAL EFFECTS

It is difficult for user to identify the object that he/she is working on in an AR view when the instructional video is overlaid. Therefore, we add various visual effects to the instructional video easily distinguished. The user carries out this processing whenever he/she wants or needs to.

5.1 Alteration of transparency degree

We carry out alpha blending processing that varies the luminance values. The alpha value is represented in Equation 3. The user can vary this alpha value so that he/she can watch an image at a time or position of his/her choice. Further, the system can distinguish among images in order to vary the alpha value in a cyclic manner.

5.2 Contour display

We have attempted to show the contours in the instructional video image in order to let the user distinguish the user's view from the video view easily. The result of showing the contours in the instructional video by using the Canny edge detector [3] is shown in Fig. 11. If the background color of the instructional video and the color of a contour are similar, it is possible to change the color of the contour or the thickness of the contour, as shown in Figs. 12 and 13.

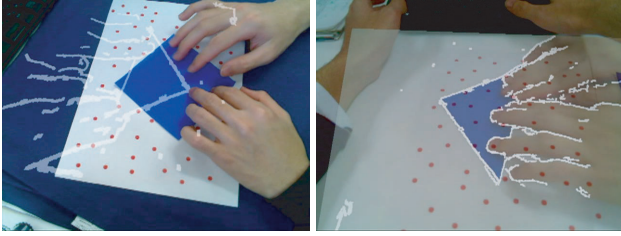


Figure 11: Synthesis results obtained by visualizing the contours of the instructional video



Figure 12: Synthesis results obtained by highlighting the contours in various colors

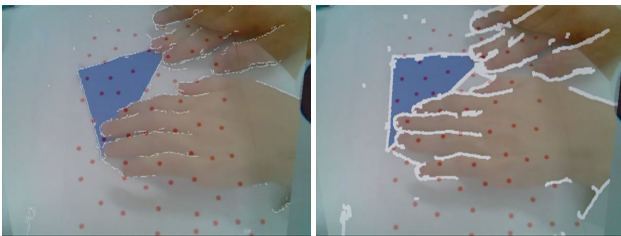


Figure 13: Synthesis results obtained by highlighting the contours with outlines of varying thicknesses

5.3 AR view shift

There is a case in which the user may want to shift the AR view completely instead of watching the AR overlaid view in a certain situation. For example, if the instructional object is changed from one conformation to another, then it is difficult to clearly see the instructions. By shifting the AR view, the user can understand the operation as if an instructor were teaching beside them. Therefore, the user will be able to distinguish between their view and the video view easily. This effect seems to be suitable for the manipulation in which the position of the object in itself is not important. Therefore, it is not suitable for the task such as involving the puzzle.



Figure 14: Overlaying results with the shift in the AR view (Left) the instructional video, (Middle) the user's view, and (Right) the shift in the AR view

5.4 Display of the new region

In the proposed system, unimportant regions such as the background of the instructional video, are just overlaid. It is believed that it is easier for the user to understand an instruction if only useful regions are displayed. Therefore, we produce an image of the new region cut off from the instructional video. Here, we regard the changing region as the new region. We use the background subtraction method for the detection of the new region, and we use the CIE $L^*a^*b^*$ color system for a robust detection under various illumination conditions. This visual effect is suitable for an instructional video for tasks such as those involving building blocks, because this effect makes it easy for the user to identify the place where he/she needs to place a block.

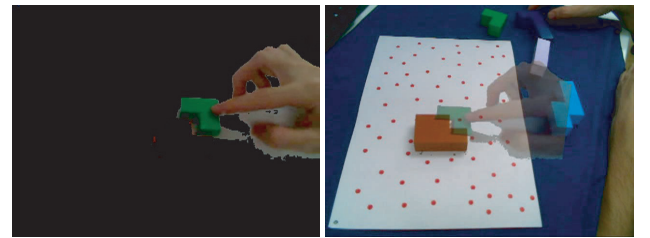


Figure 15: Synthesis result obtained by overlaying only the changed area: (Left) the changed area between the image of the instructional video and the background image, and (Right) the synthesis result

5.5 Swapping of the RGB values

We will explain the "swap RGB values" processing in order to distinguish the object being worked upon by the user from the object being worked upon in the video by extremely varying the color of the instructional video. (For example, swapping RGB with GBR or RGB with BRG) In order to swap the RGB color value of the instructional video, a distant vision needs to be created. Therefore, it is expected that this effect will solve the problem of visual confusion. The result of using this effect is shown in Fig. 16. Of course, this effect can not be used for the task which needs the color information.

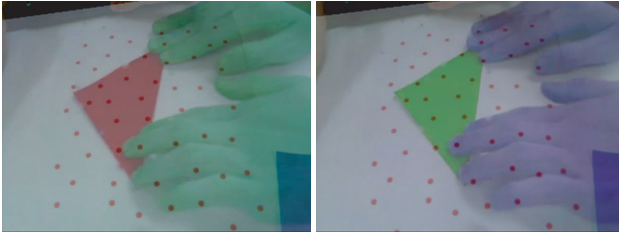


Figure 16: Synthesis results of RGB swapping: (Left) usual image, (Middle) image obtained by swapping RGB with GBR and (Right) image obtained by swapping RGB with BRG

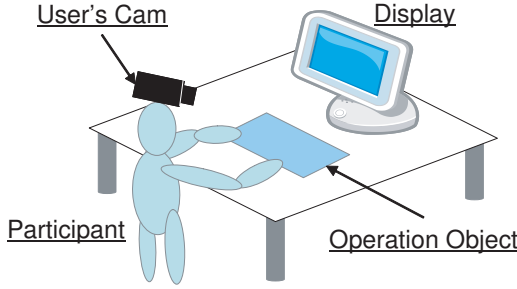


Figure 17: Experimental environment

6 EVALUATION

We have conducted user evaluation experiments for confirming various issues related to the proposed task support system. We recruited the participants from our laboratory. Because the number of participants in this user evaluation is small and the variety of the instructional videos for the evaluation is also limited, we consider that these experiments provide just preliminary results. We are planning to perform the extended evaluation experiments based on these preliminary results.

6.1 Visibility study

In this user evaluation, we confirm the significance of the two things: (1) overlaying the instructional video onto the object worked upon by the user and (2) adding various visual effects to the instruction video. The participants performed two tasks: involving building blocks and making an origami. We prepared two instructional videos in which how to form a shape by placing some colored blocks and how to make a box from a square paper are showed, respectively. These videos consist of ten steps, and all of the tasks are designed to be simple. Then, the participants evaluated the suitability of the visual effects to the tasks by using a 5-point Likert scale, from 1 to 5.

The experimental environment is shown in Fig. 17. Three participants (all males; age: 22-23 years) performed the task while watching a vision that was created by translucently overlaying the instructional video onto the user's view in the stationary display. The participants selected every visual effect and evaluated whether the selected effect was suitable for this task.

- Experimental result -

The results of this evaluation are given in Table 1.

In case of building blocks, the score of merely overlaying the display and the display of the new region were high. Because the position of the instructor's object was important in case of building blocks, we can say that the users could understand the instruction more intuitively. On the other hand, in the case of the origami task, the score of shifting the AR view and swapping the RGB values in

Table 1: 5-point Likert scale results of subjective evaluation of visual effects

5: I think the effect is suitable for this task
3: I do not know whether the effect is suitable for this task
1: I think the effect is not suitable for this task

Visual Effect	Placing Blocks	Folding Paper
Overlaid	4.3	1.7
Shift in AR View	2.7	4.0
Alteration of Transparency	3.0	2.0
Contour Display	3.3	2.3
Showing of New Region	4.0	1.0
Swapping of RGB Values	1.3	3.3

the view were high. Because the shape of the paper manipulated by the instructor should be clearly visible for making the origami box. Moreover the instructor's manual procedures are also important. Therefore, it was easier for a participant to distinguish the object that the participant was working upon from the object being worked upon in the instructional video by these visual effects.

From these results, we confirmed that the suitable visual effects were different by the contents of instructional video. In the future research, we will consider the most suitable visual effect based on these preliminary results to apply the proposed system to other tasks.

6.2 Visibility study according to the orientation of the overlaid video

The aims of this experiment is to confirm the significance of adjusting the instructional video's view to the user's view and to confirm the relationship between the visibility and the orientation angle of the overlaid video.

We prepared two different types of instructional videos (video A and video B) for the origami task. Each instructional video consists of seven different folding manipulations. The seven steps of the folding manipulations are randomly decided without any intention to generate any particular origami craft. Because we intend to avoid the effect of participants' prior knowledge and experiences for generating the particular origami craft. Therefore, they cannot predict the next procedure unless they follow the manipulation of the instructional video.

The experimental environment is the same as that of the above experiment. However, note that because we focus on only the orientation angle of the instructional video, we did not display a participant's view but displayed only the view of the instructional video whose orientation was changed. Ten people (9 male and 1 female; age: 21-24 years) participated in this experiment.

Then, we divided the participants into two groups. One group of participants watched the video A whose viewpoint was fitted according to the user's view and the video B whose viewpoint was not fitted. The other group watched those videos with adverse condition. Later, the participants evaluated the effect of the fitting of the display by using a 5-point Likert scale, from 1 to 5. The result is shown in Table 2.

Further, we measured the elapse time to complete the task by ten participants, in order to confirm whether there is a difference in the task completion speed between displaying with viewpoint fitting and without viewpoint fitting. The video comes to repeat the same step unless the user gives the sign to proceed to the next step. The results of measured elapse time and the number of repeated times are shown in Fig. 18.

Moreover, participants evaluated the effect of the every 10° rotation of the video display by using a 5-point Likert scale, from 1 to

Table 2: Result of subjective evaluation of the fitting of the instructional video

Visibility (5-point Likert scale)	number of participants
5: The fitted video is very easy to follow	6
4: The fitted video is easy to follow	3
3: The fitted video is moderately easy to follow	1
2: The fitted video is difficult to follow	0
1: The fitted video is very difficult to follow	0

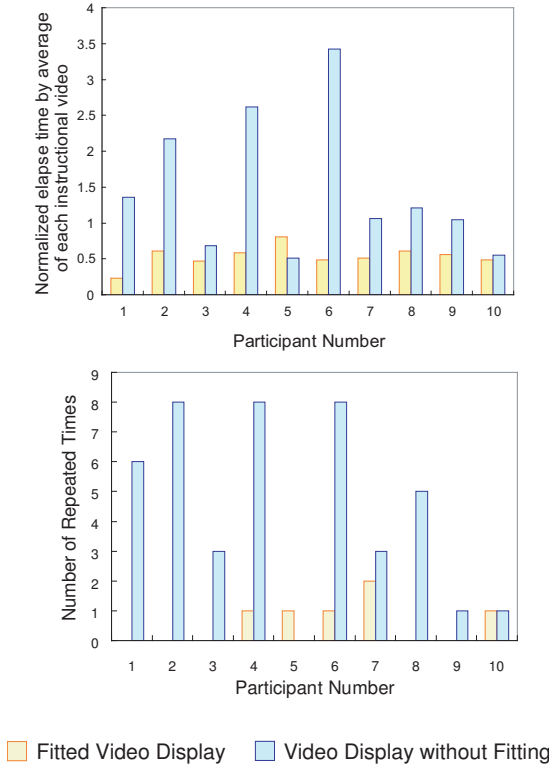


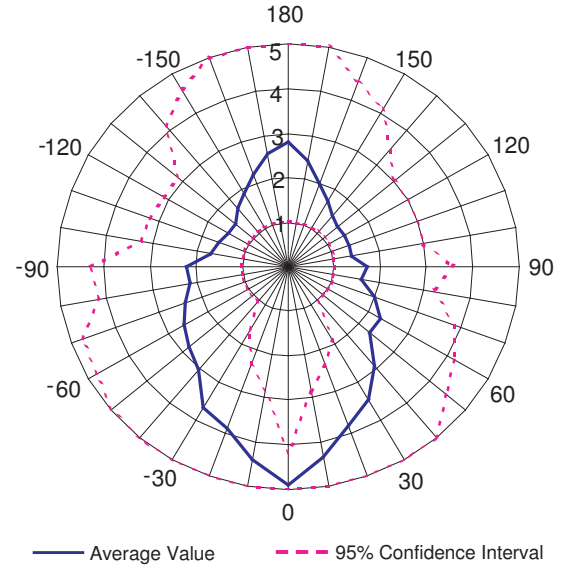
Figure 18: Working time of each subject in the cases of the fitted instructional video display and the video display without the fitting

5, in order to confirm the relationship between visibility and the orientation angle of the overlaid video. The results of this evaluation are shown in Fig. 19.

- Experimental result -

As shown in Table 2, most of the participants feel that it is easy to follow the video fitted to display according to the user's view. Similar tendency is shown in Fig. 18. 90% participants took less time to complete the task when the video was fitted according to the user's view. In case of the fitted video, all the participants finished the task in the lower time than average. On the other hand, in case of not-fitted video, some of them need to repeat the instructional video many times. The variance of the elapsed time is also big by each participant. These results confirm that the fitting of the instructional video is effective for understanding of the user.

From the results shown in Fig. 19 about the relationship between visibility and the orientation angle of the overlaid video, we observe that the range from -30° to 30° has a higher evaluation value than the other ranges. In the case of the direction opposite to the user's



5: I think the effect is suitable for this task
3: I do not know whether the effect is suitable for this task
1: I think the effect is not suitable for this task

Figure 19: Result of visibility evaluation of the effect of the rotation of the instructional video

view (180°), the variance of the evaluation value is considerable. For example, one of the participants responded that it was easy to follow the video because the image looked like a mirror reflection; another participant said that it was difficult to follow the video because it confused his sense of space. In the case of the right of the user's view (around 0°), the evaluation value is high irrespective of the participants. The range of the angles at which the evaluation value is between 3 and 5 in the 95% confidence interval is -8° to 8° . These results confirm that the proposed instructional support system is effective for almost all types of users.

7 CONCLUSION

In this study, we presented an AR-based instructional support system that uses existing instructional videos as the additional AR contents. In order to solve the visual confusion problem, we proposed the addition of various visual effects that improve the user's visibility. Furthermore, the system adjusts the pace of the instructional video pace according to the user's pace so that the user can carry on with the task at his/her own pace.

In the experiment to confirm the effectiveness of this system, we assumed two task scenes and conducted the user evaluation.

In case of the origami task, it is important ensure that the display helps the user to distinguish the object that he/she is working on from the object that is worked upon in the instructional video. The system should not completely overlay the instruction video onto the user's view space but overlay the instructional video onto the side of the user's view space and fit the instruction video to display according to the user's view.

In case of the task involving the building blocks, not only is it necessary to distinguish the object worked upon by the user from object worked upon in the instructional video, but it is also important to fit the position of the user's work space such that the user recognizes the object intuitively. We confirm that the various visual effects added to the proposed system, such as the enhancement of transparency, are useful to the user depending on the task to be per-

formed. Therefore, we can classify these effects according to the tasks to be performed and obtain a valid guideline for applying the proposed system to other tasks.

In the future, our system will eventually use a HMD, and therefore, we plan to evaluate its effects HMD.

ACKNOWLEDGEMENTS

We would like to thank Dr. Hideaki Uchiyama for providing the program of the point marker for camera tracking. He also kindly gave us helpful comments for improving our final version of the paper.

This research was partially supported by the Grant-in-Aid for Scientific Research of Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT Grant number : 21500178).

REFERENCES

- [1] "How To Videos — Grasp.com". <http://www.grasp.com/>.
- [2] "Instructables - Make, How To, and DIY". <http://www.instructables.com/>.
- [3] J. Canny. A computational approach to edge detection. *IEEE Transactions Pattern Analysis and Machine Intelligence (PAMI'87)*, 8:679–698, 1987.
- [4] M. Flagg and J. Rehg. Projector-guided painting. In *Proceedings of the 19th annual ACM symposium on User interface software and technology (UIST'06)*, pages 235–243, 2006.
- [5] S. Henderson and S. Feiner. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In *Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality (ISMAR'09)*, pages 135–144, 2009.
- [6] H. Ishii. TeamWorkStation: towards a seamless shared workspace. In *Proceedings of the ACM conference on Computer-supported cooperative work (CSCW'90)*, pages 13–26, 1990.
- [7] H. Kato and M. Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR'99)*, pages 85–94, 1999.
- [8] H. Kuzuoka. Spatial workspace collaboration: a SharedView video support system for remote collaboration capability. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI'92)*, page 540, 1992.
- [9] Y. Motokawa and H. Saito. Support system for guitar playing using augmented reality display. In *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'06)*, pages 243–244, 2006.
- [10] W. Narzt, G. Pomberger, A. Ferscha, D. Kolb, R. Muller, J. Wiegardt, H. Hortner, and C. Lindinger. Augmented reality navigation systems. *Universal Access in the Information Society (UAIS'06)*, 4(3):177–187, 2006.
- [11] S. Nilsson and B. Johansson. Acceptance of augmented reality instructions in a real work setting. In *CHI'08 extended abstracts on Human factors in computing systems*, pages 2025–2032, 2008.
- [12] D. Reiners, D. Stricker, G. Klinker, S. Muller, and Fraunhofer. Augmented reality for construction tasks: doorlock assembly. In *Proceedings of 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'98)*, 1998.
- [13] C. Robertson, B. MacIntyre, B. Walker, and G. Center. An evaluation of graphical context when the graphics are outside of the task area. In *Proceedings of 7th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'08)*, pages 73–76, 2008.
- [14] T. Salonen and J. Saaski. Dynamic and visual assembly instruction for configurable products using augmented reality techniques. *Advanced Design and Manufacture to Gain a Competitive Edge*, pages 23–32, 2008.
- [15] A. Tang, C. Owen, F. Biocca, and W. Mou. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI'03)*, pages 73–80, 2003.
- [16] J. Tang and S. Minneman. VideoDraw: a video interface for collaborative drawing. *ACM Transactions on Information Systems (TOIS'91)*, 9(2):170–184, 1991.
- [17] B. Tversky, J. Morrison, and M. Betrancourt. Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57(4):247–262, 2002.
- [18] H. Uchiyama, H. Saito, V. Nivresse, M. Servieres, and G. Moreau. AR representation system for 3D GIS based on camera pose estimation using distribution of intersections. In *Proceedings of the 18th Annual International Conference on Artificial reality and Telexistence (ICAT'08)*, pages 218–225, 2008.