

Stepping-in Presentation

Akinori YOSHINO

Hiroshi Mitsumata

Kenji MIYAMOTO

Faculty of Engineering, Hosei University,

3-7-2 Kajino-cho, Koganei, Tokyo 184-8584, Japan

TEL +81-423-87-6349

yoshino@philos.k.hosei.ac.jp mitsumata@philos.k.hosei.ac.jp miyaken@k.hosei.ac.jp

Abstract

Presentation technique in which one can give presentation from inside a slide is proposed. A presenter's live image and a slide share an area to inhabit and interact with each other. This is realized by overlaying the slide with the presenter's image from video camera and passing control signals to active contents in response to the presenter's action. As a result, the projected presenter looks as if he is standing in front of the screen and working on it. This makes, experiment shows, video presentation more natural with presenter's usability comparable to white boards. Our approach augments digital contents by overlaying them with a real image, which contrasts with Augmented Reality, in which the real world is overlaid with digital contents.

1. Introduction

We propose *Stepping-in Presentation*, a method of presentation with which one can give presentation from inside a slide projected by a computer.

Recent presentation software utilizes a presenter's video image to make the presentation more impressive. For example, Producer for PowerPoint 2002 [1] displays a presenter's video image as well as a slide. This makes stored presentation contents self-explanatory. Even in case of live presentation, they still have advantage if it is given from remote place or the size of the screen mismatches that of the presenter.

In comparison with presentation by a real person, however, presentation by video image so far have disadvantage in the following two points. The first, in presentation by video image, an image and a slide are displayed at different position in the screen as is shown in Fig. 1. This situation makes the audience embarrassed about which position to look at, the slide or the presenter's image. The latter is, presentation by video image lacks interaction between the presenter and the slide. This situation makes presenter's image isolated, limiting its expressive power. That is, if the presenter were a real person standing in front of a white board, he could draw marks, move plates with text, and so on.

In *Stepping-in Presentation*, both a slide and a presenter's image inhabit a same area and interact with

each other. More exactly, presenter's image is displayed at the position where the presenter would be seen in case of standing in front of the real screen. This cohabitation is achieved by overlaying a slide with the presenter's video image captured from a video camera. Interaction is realized by passing control signals to active contents in response to the presenter's action.

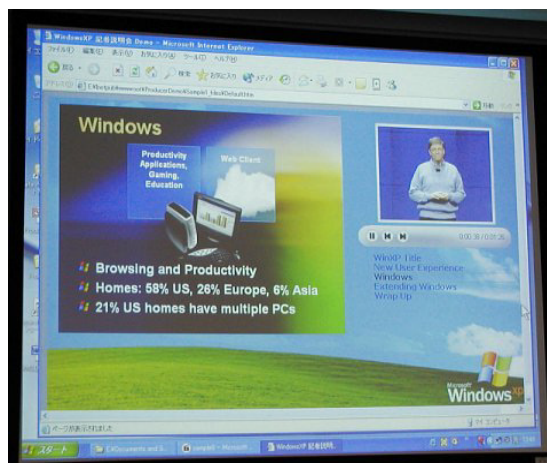


Fig. 1. Presenter and slide are separately shown ([1]).

Thanks to that the presenter and the slide are cohabiting in the same position, the audience can concentrate on watching the presenter's action. Thus, from the audience's side, the presenter seems as if he were directly pointing the slide standing in front of the real screen. That is, for the audience the presenter seems as if he "Steps in" the slide. Thanks also to interaction with objects in the same area, the presenter seems as if he were directly manipulating them just the same as using an electric white board.

Experiments show that a presenter can use the system easily. The presenter feels like presenting as if standing in front of real screen, because he can easily empathize himself's image on the screen.

Stepping-in Presentation is a cheap and easy technique for joining digital contents and video images in a natural way. In the rest of the article, after overviewing *Stepping-in Presentation*, we discuss its feasibility and usability in detail.

2. STEPPING-IN PRESENTATION

This section overviews Stepping-in Presentation, and gives its intended meaning.

2.1. Look and Feel

In Stepping-in Presentation, audience sees a presenter inhabits a slide and gives presentation from there. An actual screenshot projected on the screen is shown in Fig. 2. In Fig. 2, the image of the presenter seems to be standing in front of the screen and pointing an item on the slide, while real he is standing outside the screen. In Fig. 2, we can see buttons which the projected presenter can "press" to control the slide.

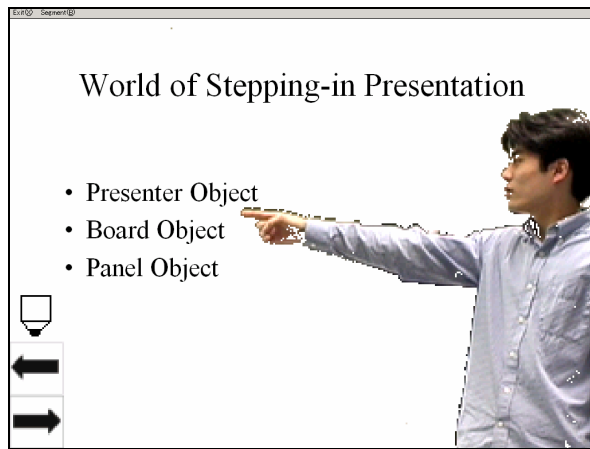


Fig. 2. Appearance from audience.

2.2. Setting up

Equipments required to give a Stepping-in Presentation are shown in Fig. 3. We need a video camera for taking a presenter's video image and a PC for both processing video data and providing slide contents. Details of the camera angle are discussed later in Sec. 3.4.

After the equipments are all set up, some calibration procedure might be needed depending on implementation. For an example, "Background Subtraction Method", which is used in our current implementation, needs to take background scene behind the presenter in this calibration phase.

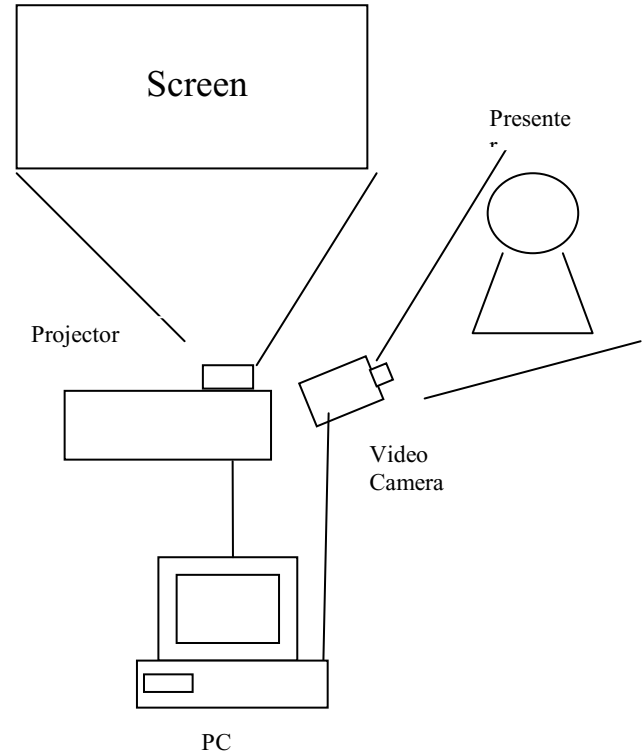


Fig. 3. Arrangement of equipments required.

2.3. Using

The presenter's actual standing point during the presentation is, as is shown in Fig3, not in front of the screen but at a different place where he can look at the screen. He does not have to care about behind him because only his image is automatically cut out from the background. He gives presentation watching the screen because actual he is not in front of the screen. If another monitor facing him is added on the audience's side, he can check the presentation while looking at the audience.

2.4. Interpretation

Stepping-in Presentation has an intended model called *Presentation Field* through which a presentation is interpreted. Stepping-in Presentation consists of objects interacting with each other. Among them are a *presenter object* as a digitized presenter, a *board object* as the original slide, *control objects* such as next-page buttons, and other objects which are imported from the real world or created by a computer. Fig. 4 illustrates the Presentation Field with these objects.

In Presentation Field, a presenter object could control all kind of objects including himself. An example of this self-control is changing visual effects of the presenter object.

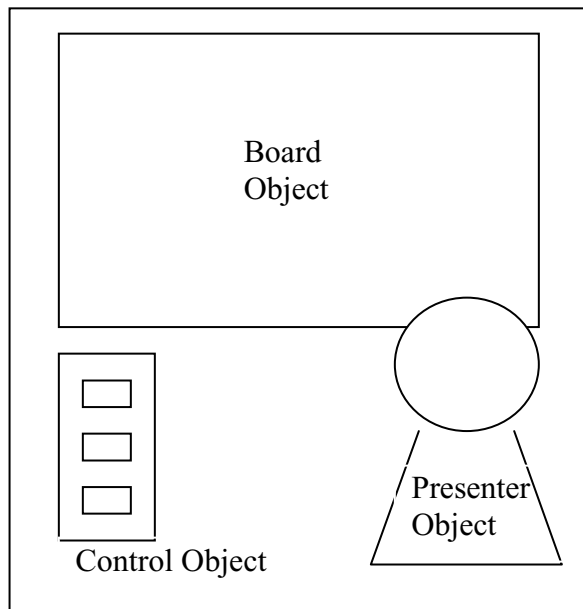


Fig. 4. Presentation Field.

3. FEATURES

This section lists up features and some usability of digitization.

3.1. Cohabitation

As we saw in Fig. 2, a presenter seems as if he is standing in front of a real screen and pointing directly an item. Thanks to this cohabitation, the presenter can give gestures in the spatial context of the slide contents displayed behind his image.

3.2. Interaction with Objects

In Presentation Field, a presenter object can interact with other object in various ways. An example of interaction with board object is writing on it with a pointing stick as a virtual pen as is shown Fig. 5. An example of interaction with a control object is pressing a button to control slide. With this functionality, he can turn the slide pages, as is shown in Fig. 6, without leaving the screen to operate the PC. In this way, the presenter gives presentation as if the he is using an electric white board. Note here that the presenter's action for interacting with other objects is visible from the audience. Thus, the presenter can make even his controlling action informative, notifying the audience his intention.

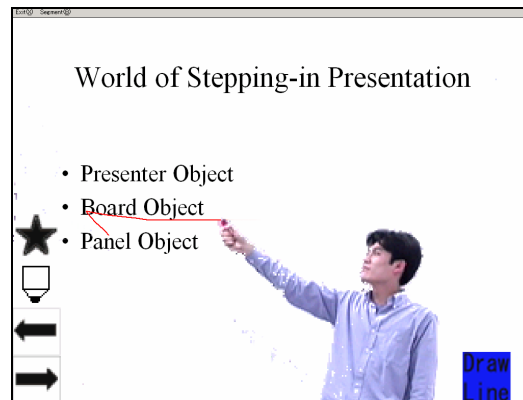


Fig. 5. The presenter is drawing a line.

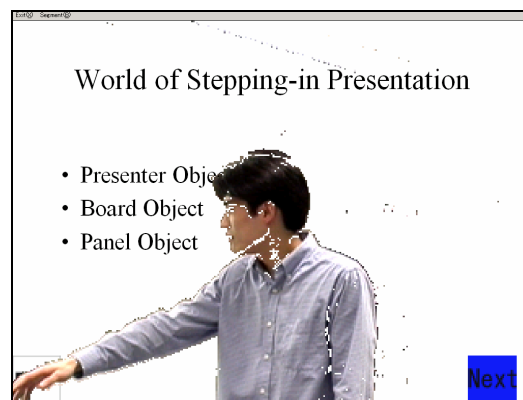


Fig. 6. The presenter is pressing a button object.

3.3. Advantages of Digitizing Presenter

Expressive power beyond white boards comes from a presenter being separately digitized. Once a presenter is digitized apart from other slide contents, one can choose from various operations in putting it and the other contents together.

First, before overlaying the slide, the digitized image can variously edited for adding visual effects. For instance, a presenter's image can be made transparent as is shown in Fig. 7. Relocated or rescaled it can be in the same way. Note that, these operations can also be commanded by the presenter during the session. In other words, a presenter can conduct stage effect by himself.

Second, in composing the whole scene, presenter's image can be placed in various depths relative to other objects. In Fig. 8, presenter's image stands behind the molecule model. Whereas in Fig. 9, the presenter's image stands in front of it.

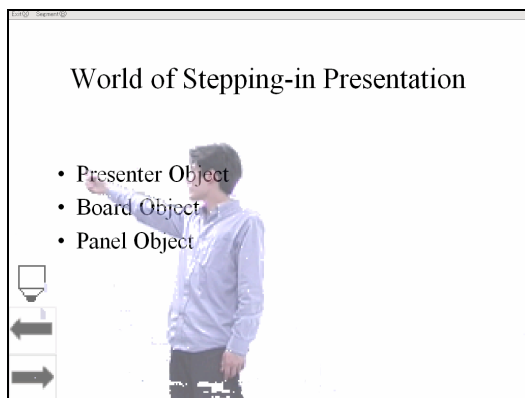


Fig. 7. The presenter was changed transparent.

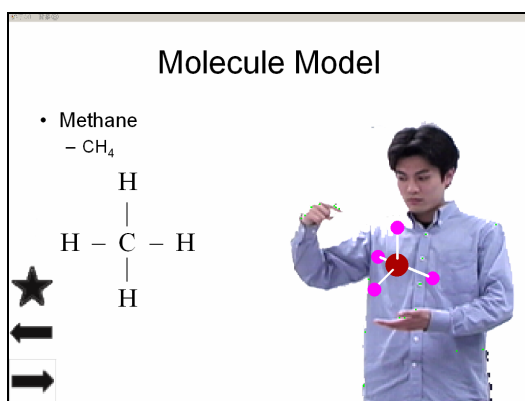


Fig. 8. The presenter is standing behind the molecule model.

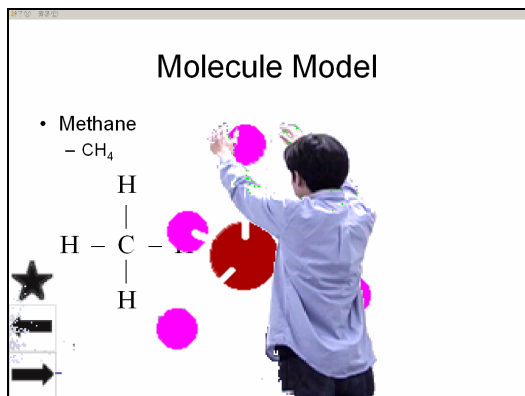


Fig. 9. The presenter is standing in front of the molecule model.

4. USABILITY EVALUATION

In this section, after seeing how the appearance of the presenter's image changes according to the variation of the camera angle, usability evaluation with each angle is presented. Usability here is that from audience's side and that from presenter's side. Experiment was performed

using a PC with Windows 2000, Intel Pentium4 processor 1.4GHz, 256MB memory, NVIDIA GeForce2 GT video card, and InsideTNC TV WDM Video Capture, with which connected via S-Video Panasonic NV-MX3000 video camera, taking 30 frames per second.

4.1. Changing Camera Angle

Geometrical relations between equipments and a presenter are shown in Fig. 10. In Fig. 10, the imaginary presenter who seems to be standing in front of the screen is shown by a dotted circle. γ is the angle of the presenter's eyes against the screen. θ is the camera angle against the screen. ϕ is the angle of the presenter's eyes against the imaginary board.

The point which the projected presenter seems to look at should preferably coincide with the point which the real presenter looks at. In Fig. 10, this point is indicated by A . In this case we have

$$\theta + \gamma - \phi = 90^\circ. \quad (*)$$

Note that this holds only at certain point A on the screen. We assume that places to be pointed are all around such point and therefore the formula approximately holds in what follows.

In order for the projected presenter's finger to point at A , his real finger should be on the line BV . In addition to this condition, in order that the direction at which the presenter looks and the direction at which he points coincide, his real fingertip should be located at point B .

We examined how usability changes according to θ changing as $\theta = 0^\circ, 60^\circ$, and 90° . Other conditions were fixed as follows: Value of γ to 45° , distance between the presenter and the screen to 3.5 meter. The size of screen was 2.5 meters by height and 2.0 meters by width.

Camera positions with θ of above values are indicated in Fig. 11. For each value of θ , the value of ϕ is determined to correspond $-45^\circ, 15^\circ$, and 45° using the equation (*). Using these ϕ , positions of imaginary presenter were plotted as is shown in Fig. 11.

Resulting screen view and whole scene viewed from the audience are shown in Figs. 12-14. In these Figs, the left is the screenshot of the presentation with the presenter facing to the screen and the right is the view from the audience with the presenter facing to the camera.

Actual evaluation of these camera angles is given in the followings.

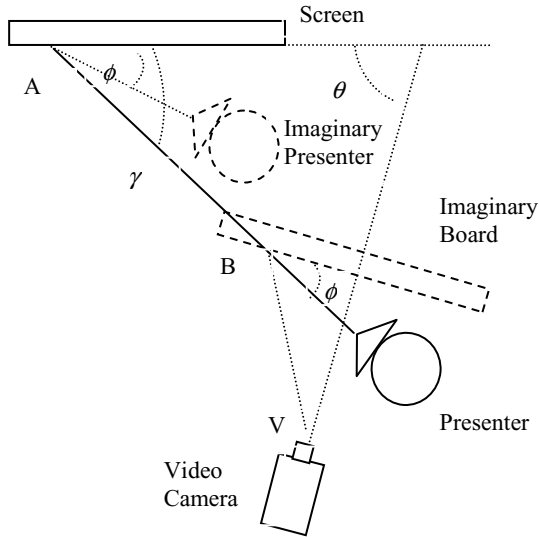


Fig. 10. Geometric relation between equipments.

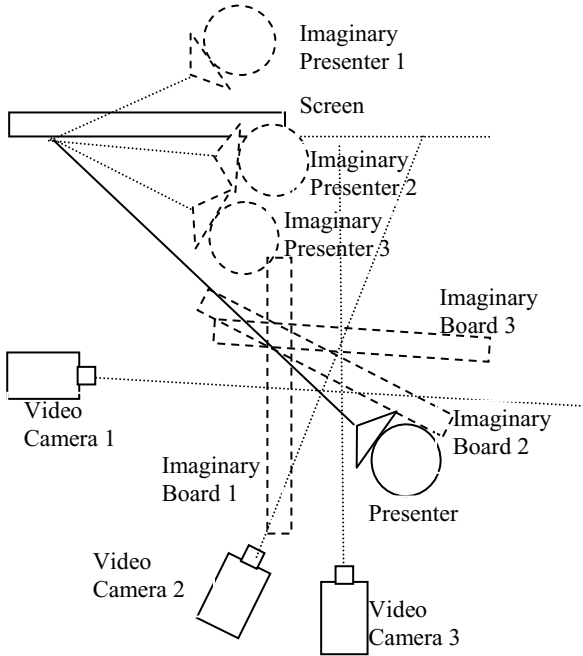


Fig. 11. Camera angles used in experiments.



Fig. 12. Screenshot and view from audience with $\theta=0^\circ$.



Fig. 13. Screenshot and view from audience with $\theta=60^\circ$.

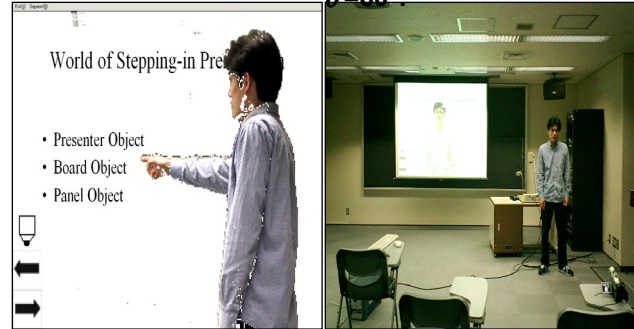


Fig. 14. Screenshot and view from audience with $\theta=90^\circ$.

4.2. Evaluation from Audience's Side

6 subjects (male students) were given a real lecture in 3 different methods, (A) Stepping-in Presentation, (B) presentation using a laser pointer, and (C) presentation by a real person standing in front of the screen. The Subjects were to compare them and were to answer the question "What was the preferability order?"

Results were that 4 of 6 answered (C) > (A) > (B) in preferability order ((C) is the best). The rest 2 answered Stepping-in Presentation is the worst. The former 4 subjects commented that (A) was good because it was easy to find where the presenter was pointing at, the latter 2 subjects commented that the presenter interfered the slide by covering it.

The same subjects were to compare the camera angles 0° , 60° , and 90° . They were to answer the question, "Which camera angle made the presenter's image natural?" Results are shown in table 1 and plotted in Fig.15. In table 1, number n in each row for Subject Number means the n -th position in naturalness order. Score is the sum of points which are obtained by converting the positions 1, 2, and 3 to points 5, 3, and 1. The result is that 60° is the best camera angle for audience.

Table 1. Naturality score for angles.

Subject No.	Camera Angle		
	0°	60°	90°
1	2	1	3
2	2	1	3
3	2	1	3
4	1	3	2
5	3	2	1
6	3	2	1
Score	16	22	15

4.3. Evaluation from Presenter's side

The same subjects as above were also to evaluate Stepping-in Presentation from presenter's side. The subjects were to perform two kinds of tasks, pointing ten letters and tracing the contour of a rectangle, under camera angles (a) 0°, (b) 60°, and (c) 90°. They were evaluated and scored by the authors with respect to accuracy and speed.

The score summed over the subjects is shown in Fig.15. Results were that for all the subjects (c) was the easiest. Under other two conditions, they were poorly-skilled for the first time. However, they all got skilled in few seconds under all conditions.

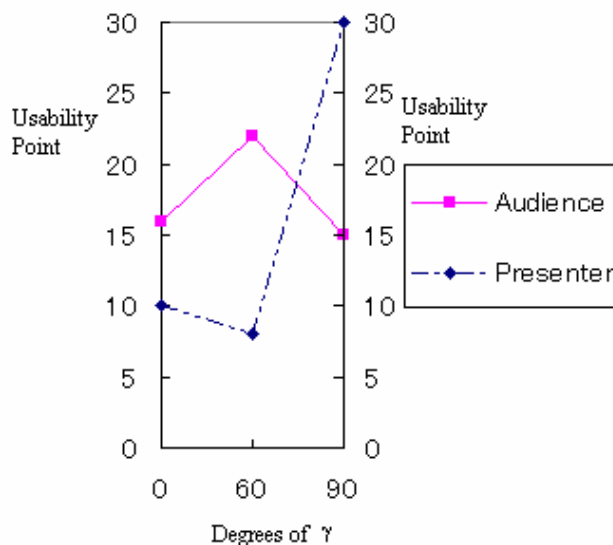


Fig. 15. Audience's naturality impression and usability for presenter.

5. IMPLEMENTATION AND STRUCTURE OF THE SYSTEM

The current implementation of Stepping-in Presentation is built up from the following modules:

1. SEGMENTER cuts out only an image of presenter from a raw video image,
2. DETECTOR detects presenter's action and translates them to commands,
3. DISPATCHER dispatches each command to other module,
4. IMAGE EDITOR adds visual effect to the presenter's image, and
5. COMPOSER puts all images together.

These Modules and data flow among them is shown in Fig. 16.

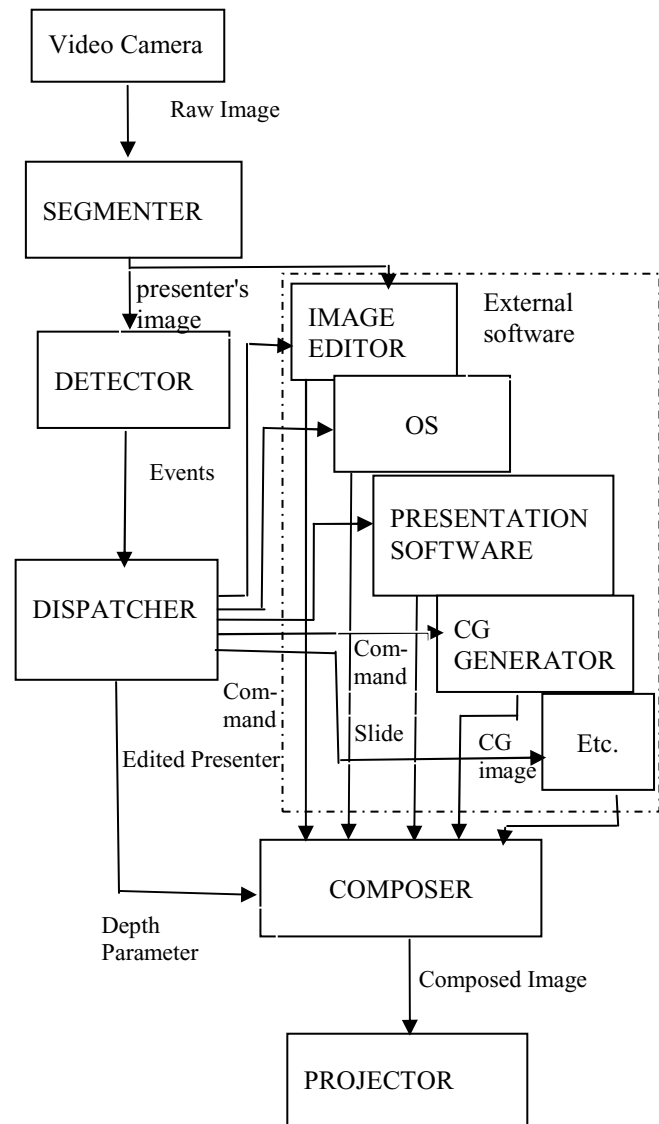


Fig. 16. Modules and Dataflow among them.

First, the SEGMENTER segments the presenter's image from the live video image by "Background Subtraction Method". In this method, the captured video

image is subtracted by the background scene preliminary taken. For each pixel i , d_i is difference in RGB at i between the background scene and the live presenter's image*. Here a color pixel has 24bit RGP. When the d_i is smaller than threshold value T , which is set to 30 in this implementation, the SEGMENTER changes the color of pixel i to specific color, which is GREEN=RGB(0,255,0) currently. The problem here is that presenter's shadows change the intensity of background, which causes undesirable region to be cut out. To avoid this, current SEGMENTER implementation does not cut out regions in which RGB is proportionally changing.

Secondly, the DETECTOR detects the presenter's specific action and produces commands. For this purpose, certain conditions are checked in each frame. An example of such conditions of presenter's image is whether a certain region has been covered by the presenter's image for more than 60 frames (which corresponds to 2 seconds by usual DV camera at 30 fps). The result is used to determine if a button is selected. Currently available information is about if a certain region is covered or about where an object of specific color is.

Thirdly, the DISPATCHER dispatches each command to COMPOSER and external software modules such as the IMAGE EDITOR, Presentation Software, CG Generator, and so on. These modules can be added easily because they are external. In addition, the presenter can launch an application through OS.

Next, the IMAGE EDITOR adds visual effect to the presenter's image as was described in Sec. 3.3.

Finally, the COMPOSER puts together fragments of image incoming from the above software modules according to the occluding relation determined by their depth parameter. GREEN part of the image coming from the SEGMENTER is made transparent here. The COMPOSER makes the slide be overlaid with the presenter's image, the CG image, and so on.

In current implementation, a method of presenter's interaction is simplified. For example, if a certain region is kept covered for same period, the system interprets it as the region is clicked.

The prototype system was developed in Microsoft Visual C++ 6.0 and DirectX ver. 8.

6. DISCUSSION

We proposed Stepping-in Presentation, a method of presentation with which the presenter can give presentation from inside a slide projected by a computer.

From audience's side, projected presenter seems as if he is standing in front of the screen. This situation makes

Stepping-in presentation quite similar to presentation that uses a white board or something like that. Therefore, the audience feels it quite natural.

From the presenter's side, on the other hand, one may think it difficult for a presenter to let the projected image of himself act as he want. Experiments, however, found it very easy for a presenter to handle his own live image.

Stepping-in Presentation is compared with presentation using electric white board capable of displaying a slide. Both allow a presenter to interact with digital contents on a slide. Moreover, if video image of the presenter standing in front of an electric white board is taken and projected on another screen, cohabitation similar to Stepping-in Presentation can be realized. Whereas, an advantage of Stepping-in Presentation is, as described in Sec3.3, that separately digitized image can be easily handled in mixing with original slide. If it is used as stored self-explanatory presentation, Stepping-in Presentation can be regarded as annotated slide in which the video annotation can easily be turned off. On the other hand, in video of a presenter in front of the electric white board, the original slide is transferred to low-level video data and mixed with the presenter's image causing the original slide unable to be recovered.

Weather broadcasts have used the technique of merging presenter's image and slide. The weather broadcasts, however, has less naturality because the point which the presenter looks at does not coincide with the point which he directs, whereas in Sec. 3.4 our approach solves this problems.

6.1. Applicability and Limitations

The presenter's power to manipulate objects on a slide is limited by the power of action recognizer. For example, a manipulation in 3D world is possible with recognizer capable of detecting depth.

The digitized appearance of a presenter, technologically, can be modified arbitrary. However, modification tractable for a presenter is limited by human cognitive power. For example, presenter rotated 90° would be difficult to handle.

6.2. Related Works

In Virtual Reality software, avatars are used for a user to step in a virtual world. In presentation, however, presenter's live expression plays important role. Therefore, our approach has advantage at this point.

Our approach can also be regarded as a special case of CSCW (Computer Supported Cooperative Work). ClearBoard [2] also underlays the workspace with his collaborator without being segmented. Here, his own image is not superimposed on his workspace. In Stepping-in Presentation, more than one presenter, possibly at different places, could collaborate.

* Difference in RGB at pixel i between images P and Q is defined as $\max\{|r_i(P) - r_i(Q)|, |g_i(P) - g_i(Q)|, |b_i(P) - b_i(Q)|\}$, where $r_i(P)$ is intensity of red color at pixel i of image P and so on.

Let's compare Stepping-in Presentation with Virtual Reality or Augmented Reality [3]. Characteristics of the three are shown in table 2. Virtual Reality and Augmented Reality aim at their world to be realistic. Stepping-in Presentation does not care about the world to be realistic and nonrealistic. If his pointing agrees with his eyes, the world becomes realistic. The presenter's image could also be modified to be super natural.

6.3. Conclusion

1. We proposed Stepping-in Presentation, in which a presenter steps into among digital contents with his facing and pointing consistent, and
2. the presenter can even interact with these digital contents, making the presentation more expressive. Moreover,
3. the system is easy to use at anywhere and its cost is minimal (a PC and a video camera).

6.4. Future Work

As future work, other presenter at remote place should be enabled to step in presentation by combining multiple systems with shared image as is shown Fig. 17. This completes bidirectional CSCW.

What kind of interaction is effective in Stepping-in Presentation should be investigated further.

ACKNOWLEDGMENTS

The authors thank to Mr. Humihiko MACHIDA (student at Hosei Univ. at that time) for his many useful comments for implementation and reviewers of ACM UIST 2003 for suggesting usability experiment.

REFERENCES

- [1] Producer for PowerPoint, <http://www.microsoft.com/office/powerpoint/producer/default.asp>, Microsoft Corporation (2002)
- [2] Ishii, H. and Kobayashi, M., "ClearBoard: A Seamless Media for Shared Drawing and Conversation with Eye-Contact," *Proceedings of Conference on Human Factors in Computing Systems (CHI '92)*, ACM SIGCHI, Monterey, 3-7 May 1992, pp. 525-532.
- [3] Rekimoto, J., Nagao, k., "The World through the Computer : Computer Augmented Interaction with Real World Environments", *Proceedings of UIST '95, ACM Symposium on user Interface Software and Technology*, November 1995, pp. 29-36.

Table 2. Comparison of Virtual Reality, Augment Reality, and Stepping-in Presentation

	Virtual Reality	Augmented Reality	Stepping-in Presentation
Place of World	Computer	Real World	Computer
Characteristic of World	Real	Real overlaid with virtual	Virtual overlaid with Real

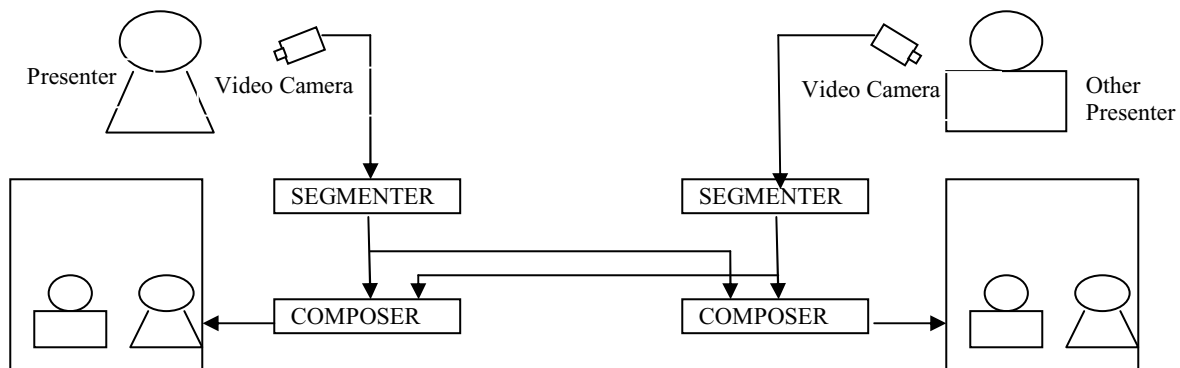


Fig. 17. Stepping-in Presentation with remote two presenters.