Real-time Reconstruction of Pseudo Wide-angle Images with an Approximating Depth Model

Kenji.Honda, Naoki.Hashimoto and Makoto. Sato

Precision and Intelligence Laboratory, Tokyo Institute of Technology, Japan
E-mail : {khonda, naoki}@hi.pi.titech.ac.jp, msato@pi.titech.ac.jp

Abstract

Recently, interactive contents with high quality images are widely used in various areas. One of the ways that makes the contents more realistic is displaying them as covering viewer's field-of-view. However, the view-angle of the existent contents are not enough for covering the large display surfaces because they are based on household TV monitors whose target view-angle is about 30 degrees. In order to achieve realistic feeling of being in that content, we have to enlarge the view-angle of these images in real-time. So, in this research, we propose a technique that reconstructs wide view-angle images by extracting peripheral images from time-series image frames in real-time. A simple 3-D model approximating target scenes contributes to, achieving real-time processing and reduction of uncomfortable feeling with distortions of the reconstructed peripheral images. We apply this method to an actual game content, and consider the quality of image and its processing time. By using that implementation, we perform objective and subjective evaluations, and confirm the effectiveness for enhancing the fun of the content itself.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation I.4.8 [Image Processing and Computer Vision]: Scene Analysis

1. Introduction

Recently, interactive contents with high quality images are widely used in various areas. Such as virtual reality (VR), computer-generated images used for some walk-through in virtual cities have enough quality to be seen as an actual scenery. And also, in recent video game consoles like Sony PlayStation3 [SCE] and Microsoft Xbox360 [MS], every users can enjoy high quality images as if they were actually in virtual worlds represented with high quality images.

One of the ways that makes these contents more realistic and gives high immersion to the users is displaying them as covering viewer's field of view. The recent spread of large size TVs in homes is caused by having feeling of being there by displaying high quality images on the large screen. That innovation will continue to familiarizing super hi-vision [JBC] which has 100 degree view-angle as a next generation TV system. As for VR applications, immersive projection displays like a CAVE system [CNSD93], have been actively developed, and used as a de facto standard for human-scale virtual environments.

On the other hand, the image contents for the wide view-angle displays are generated and used only for research, amusement and some exhibitions. Although household image contents have enough quality to be used in such wide-angle displays, their view-angle is still limited to the conventional TV size whose view-angle is about 30 degree. This condition will continue well into the next decade. Practically, by using projectors, we can easily enlarge the image contents. However, it leads to the lack of resolution and uncomfortable feeling with its unnatural scale. So we cannot fully enjoy the contents in such approaches.

Therefore, in this research, we propose a new technique to easily display the accessible contents including interactive games on the wide view angle displays with real-time processings. This method achieves pseudo expansion of the view-angle of the contents, and aims to enhance the fun and immersion with the large display systems.

2. Related works

Some techniques to increase the view-angle of the images have been proposed in the fields of computer graphics(CG),

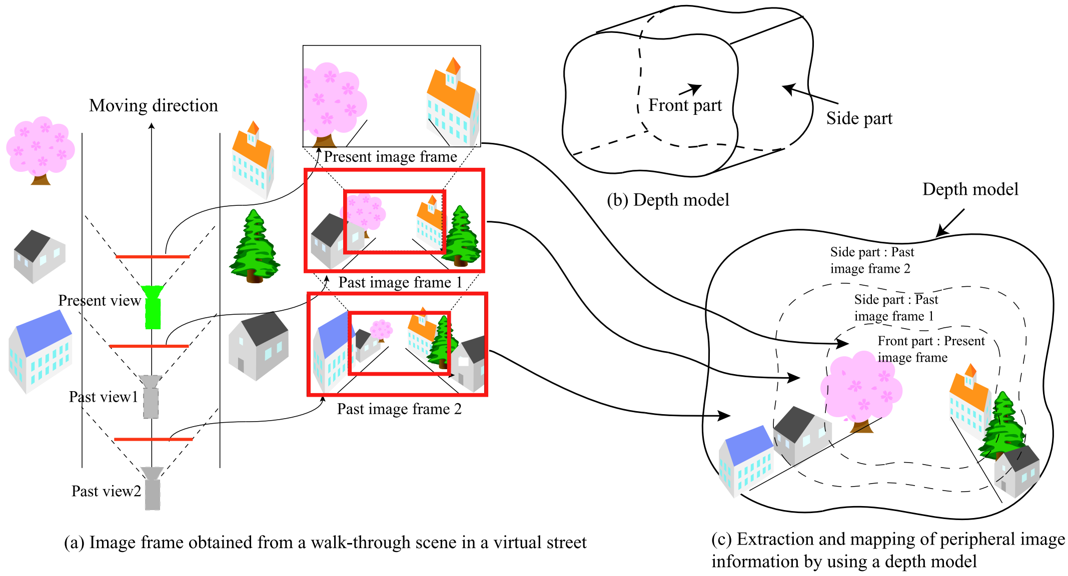


Figure 1: *Extraction of peripheral image information by using an approximating depth model*

computer vision (CV) and VR. WireGL [HEB*01] and Chromium [HEB*02] can increase the view-angle of the existing 3D-CG applications, and display their output images on the large displays consisting of multiple PCs. These techniques intercept OpenGL's instructions and distribute them over the multiple PCs through a network. And also, self-distributing software environment [HIS06] can generate wide view-angle images with the game contents running on the multiple PCs. The process order of the PCs are well synchronized, and the camera parameters for generating images are properly modified to cover whole screen with the total rendering region with the PCs. However, these techniques intend the image that is dynamically generated with PC's applications. They use the way that the applications use dynamic-link mechanism to make adjustments necessary to the expansion of field-of-view. Because such situation is a rare case, we cannot support our accessible image contents, like DVDs and TV games.

In order to support various contents, we have to expand them only with image frames. Video mosaicing [ISI*06] is one of the approach to reconstruct peripheral images with time-sequential image frames, not using special characteristics as mentioned above. However, the video mosaicing requires the estimation of highly accurate 3-D models of target scenes and camera parameters to combine exactly the images extracted from the time-series frames. That is to say, this approach is not suitable for interactive contents whose scene changes at every moment like TV games, because the estimation is too time consuming.

On the other hand, video mosaicing techniques without the 3-D models are also proposed to avoid the detailed estimation [Sze94]. These techniques easily estimate the parameter of the projective transformation by assuming the 3-D scene to be a simple plane displayed on a screen. They can generate wide-angle images in real-time by using no accurate estimating process of the 3-D shape of the scene. These techniques are used in case that the viewer looks all round in staying on-site. However, if these techniques are applied to some actual contents like a walk through in a 3-D space, the distortions trigger discontinuous optical-flow in peripheral view generated with these techniques. It also invokes uncomfortable feeling and motion sickness. These negative effects seriously damage the original fun of the contents.

3. Pseudo expansion of field-of-view for interactive contents

3.1. Concept

As mentioned in the preceding section, extracting peripheral information surrounding present view from the past image frames is one of the methods to expand the field-of-view of various image contents. In addition, generating peripheral images without uncomfortable feeling requires depth information of the target scene. The depth information must be obtained in real-time for supporting interactive contents. So, in this research, we introduce a simple depth model based on primitive shapes such as a rectangular solid and cylinder. By estimating limited depth informations in real-time, the shape of the simple depth model is adapted to an actual scene for achieving accurate approximation.

However, in this approach, the distortion and discontinuity of the generated peripheral images are not entirely re-

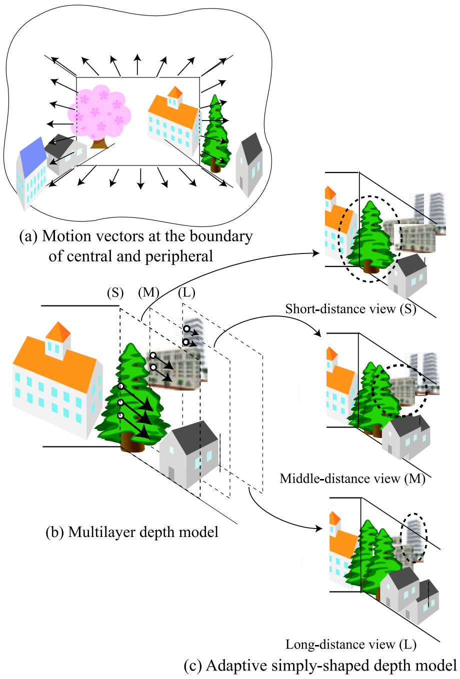


Figure 2: *Adaptation for dynamic scene*

solved, because the simple depth model is merely approximating real scene, not precisely representing that.

Therefore, we dynamically change the size of the simple depth model based on the depth information obtained from well-observed region in the target scene by heuristics or head-tracking . It leads to reducing the uncomfortable feelings with the distortions and discontinuities of the generated images. The other regions are not effective for precise recognition because of the user's perceptual characteristics. This approach also contributes to achieving real-time processing with reduced target for calculating depth. In this research, we call this method a pseudo expansion of field-of-view.

3.2. Extraction of peripheral images with a simple depth model

Figure 1 shows an overview of the reconstructing peripheral images with past image frames.

Figure 1(a) illustrates a walk-through scene in a virtual street of which both sides are surrounded with buildings, trees and etc. In this figure, the outside of the present image frame is included in past image frames. Our proposed technique achieves the pseudo expansion of field-of-view by extracting the images from these past image frames. In order to extract the peripheral images, we use the simple depth model.

At first, we transform the depth model to approximate the target scene by using motion vectors between the image frames. The details of this process are described in the next section. In Figure 1(b), the depth model is composed of a front part that covers the region of central vision of viewers and a side part that covers the region of the peripheral vision. As shown in Figure 1(c), peripheral images are extracted with the depth model. The extracted images are treated as textures, and mapped to the depth model. The front part is mainly covered with the present image frame, and the side part is covered with the images extracted from past image frames. And finally, by rendering the depth model from viewer's viewpoint, we can generate pseudo wide field-of-view images

3.3. Dynamic adaptation for changing scenes

When the depth model is inadequate to approximate the depth of actual scenes, the distortions will occur within the extracted images. These distortions have almost no effects on viewers because they occur in peripheral view. However, we can perceive distinct boundaries of the extracted images at the part illustrated as broken lines in Figure 1(c). They have strong effects because the obvious discontinuities cause sudden change of optical-flows in the peripheral view which is well-suited to perceive such stimuli. In order to reduce the discontinuities, we transform the depth model to adapting to the particular area where the viewer's often focus on. Figure 2(a) shows the present image frame illustrated in Figure 1. In this section, for ease of explanation, we generate peripheral images with a single past image frame.

As mentioned above, in the generated images, the most discontinuous area is the the boundary of the present image frame and the other part. The motion vectors at this boundary by using block-matching are illustrated as arrowed lines in Figure 2(a). Because the discontinuity at this boundary area causes intensive uncomfortable feeling, the model is estimated again by calculating the depth from this motion vectors. Although a highly accurate depth model is able to be estimated with all of the motion vectors, the calculation of the motion vectors and the estimation process require lots of time as obstructing real-time processing. In order to reduce the calculation and estimation processes, we introduce a primitive-based depth model approximating the target scenes.

3-D scenes are often treated as a short-distance view, a middle-distance view and a long-distance view according to the distance from a view-point. Figure 2(b) shows the situation when the depth model is constructed with this approximation. However, while the estimation processes undergo a simplification in this model, the extraction from past image frames on the each distance level increases extra processes like a segmentation and combination based on the distance.

Accordingly, our proposed technique uses only one of the approximated layer shown in Figure 2(b) as a representation of the whole scene. Then, in order to reduce the uncomfortable feelings, we eliminate the discontinuities by dynamically adapting the depth model to the distance obtained from the motion vector as shown in Figure 2(a). Figure 2(c) shows three situations that, for example, the depth model adapts to the short-distance view, the middle-distance view and the long-distance view. In these cases, as shown by broken lines in Figure 2(c), we can find that the object at the distance to which the depth model adapts is presented without discontinuity. Although some discontinuities occur with the mismatch of the depth model and the target area, we cannot perceive that because they are away from our concentrating region. Practically, we calculate the depth of the region where the viewers often focus on, and the depth model adapts to the distance in real-time.

Furthermore, extracting the peripheral images requires the estimation of the camera parameters like a translation, rotation and zoom. In case of expanding the contents moving forward in the 3-D space, peripheral images can be generated from a little previous frame toward the present image frame. So, we presume that the camera parameters do not change between the past frame and the present frame. And also, in our proposed technique, the movement of the camera is considered as the movement of the scene itself. Actual camera movement is represented as the transformation of the depth model mentioned above. That is to say, we can continuously use initially estimated camera position. In case of moving slowly in the 3-D space, the distance between the past image frame and the present image frame tends to increase. Although ,in that case, the peripheral images are distorted with changed camera parameters, they have little effects on the users because they are in the peripheral view and no discontinuities of the optical-flow occur. In case that the distance between the image frames is further increased, the distortions can be also reduced by using multiple past frames for extracting peripheral images.

4. Implementation

In this section, we implement the pseudo expansion of field-of-view on an immersive projection displays composed of a PC cluster. We also consider image quality and process time of the peripheral images generated with this technique.

4.1. Expansion of field-of-view on PC cluster

We implemented our proposal method on immersive projection displays "D-vision" [HHS04]. D-vision has a large screen with a full 180 degrees field-of-view. The screen is divided into 24 regions, and each region has a PC and a projector for image generation and projection. In this implementation, we used game contents with a game console as input images, and established large image projection by expanding the view-angle with 24 PCs. This architecture of this system is shown in Figure 3.

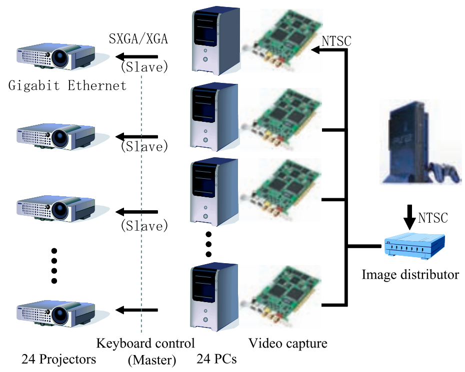


Figure 3: *Expansion of field-of-view on multi-projector displays*

In order to develop applications on a PC cluster, rendering streams and images are often distributed through the network connecting the PCs. This strategy tends to become a bottleneck of total performance because of the limited bandwidth of the network. So, in this research, we performed minimum communication to synchronize each of the PCs. The distributed images are inputted to each PC, and proposed technique expands its view-angle. When the generated image is presented, we use the projective parameter adjusted to the rendering area of each PC. This will present the image with which the whole screen of D-vision is covered. This system can provide high performance because it requires little communication-process between PCs. Although the system also distributes keyboard-events trough the network to control PCs, it little effects on the performance of applications.

We used driving-game's images of which format is general NTSC (480i). In order to prioritize processing speed, we performed the pseudo expansion of field-of-view with a rectangular solid shape as a depth model which was the simplest model of the approximation of 3-D scenes. Peripheral images were also extracted from the present image frame and single past image frame for high speed processing.

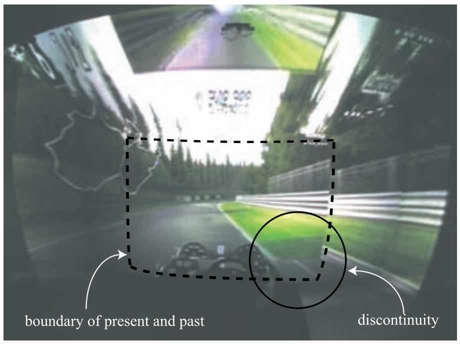
4.2. Image quality and process time

Figure 4(a) shows an original image projected to the screen as covering 60 degrees of user's field-of-view. Figure 4(b) shows the expanded image with our proposed method. The central part of Figure 4(b) is equal to that of Figure 4(a), and the peripheral part of Figure 4(b) is generated from a past image frame. If the depth model of our proposed method was not adapted to the scene, some discontinuities occurred at the boundary of present and past frame in Figure 5(a). Compared with this result, the adaptive depth model could generate less distorted and discontinued peripheral images as shown in Figure 5(b). The depth model was adapted based on the region where the users often focused on. For example in Figure 5(b), driving-road and road-side objects like a guardrail were seamlessly connected. Although some dis-

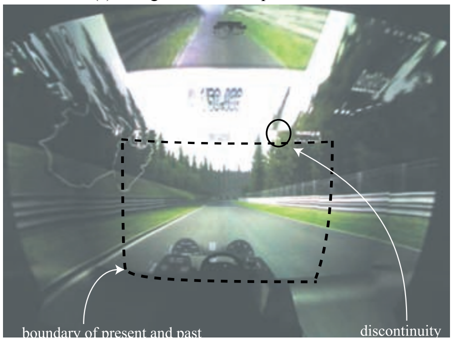
(a) Original image projected to D-vision

(b) Expanded image with our proposed technique

Figure 4: *Expanded image projection on immersive projection display "D-vision"*



(a) Using the normal depth model



(b) Using the adaptive depth model

Figure 5: *Effect of adaptive depth model*

continuities caused with the mismatch of the depth model and the target scene were still remained as shown in Figure 5(b), the viewers were not aware of them because they appeared at their peripheral view.

As for the quality of the generated images, the resolution of peripheral regions generated with a single past image frame was lower than that of the central region. This tendency is almost no problem for peripheral image generation because the peripheral view of user's are not so sensitive to the image qualities. If high resolution images are required to the peripheral, using multiple past image frame can resolve this request.

We also considered the processing time in this implementation. The simple depth model contributed to keeping 30 fps. These results were obtained with a PC which has PentiumD 3.2GHz processor, 1 GB memory and GeForce7800GTX GPU. The template size of the block-matching used for the search area of the motion vector was 8×8 pixels, and the range of the search was neighborhood 30 pixels. It is quite important to maintain the update rate (30fps) for keeping smoothness of images and response of interactive contents. Especially for the game contents, a delay of only 1 frame can easily diminish amusing factors. In this implementation, we could enjoy the game contents without any uncomfortable feelings.

5. Experimental results

In this section, we perform some evaluations of the implemented system described in section 4. By using actual game contents, we discuss whether the system can contributes to enhance reality and fun of the contents.

In this experiment, we used the driving game content shown in section 4. In the contents, some buildings are arranged at different distance from the viewpoint as surrounding users. In the super hi-vision project [EMSN07], 100 degrees of the view-angle is used to achieve enough immersion. So, in this experiment, we used 120 degrees of the view-angle for image projection.

In order to compare the effectiveness of our proposed method, we used 3 different kinds of method to project images on the large screen covering 120 degrees of the view-angle. In the first method, original images were evenly expanded to fill the field-of-view of the viewers. This style is commonly used in our home. In the second method, we expanded the images with the video-mosaicing method, as describe in section 2, which assumes the shape of target scenes as a plane. In this experiment, we treated this method as a conventional approach. And in the third method, we applied our proposed method. The central 60 degrees of the view-angle was covered with a present image frame, and the peripheral regions were filled with the images extracted from past image frames.

For subjective evaluations of immersion, we used ques-

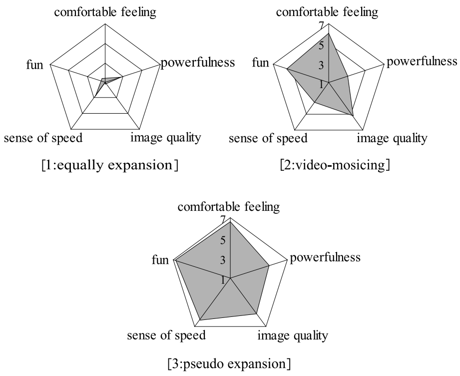


Figure 6: *Results of subjective evaluation*

tionnaires. Subjects played the games with a special seat and a handle-typed controller as shown in Figure 4. After they enjoyed the game, they were instructed to answer some questionnaires. We had a total of 5 questionnaires about comfortableness and powerfulness which were strongly correlated with immersion, image quality and sense of speed which depended on image generation methods, and fun of the contents. Each questionnaire was rated on a scale of one to seven. The results are shown in Figure 6.

In this figure, the third method with our proposed technique marks higher score than the second method with conventional video-mosaicing. In almost items except image quality, the effect of our proposed technique is significant at the 0.05 level of significance with the wilcoxon test. Discussing the results in more detail, expanding field-of-view with second and third method are significantly effective for almost items. Only expanding the image itself loses image quality and appropriate optical-flows. As a result, it leads to lose the reality of the presented images. The expansion of the filed-of-view in second and third methods resolve these problems. Especially in the third method, the optical flows in the peripheral areas are well-simulated, and the subjects can feel correct sense of speed.

Finally, from the result of the fun of the content itself, our proposed technique achieved highest points in the three methods. The purpose of expanding the view-angle of the contents is to increase and extract increase the potential fun of the contents. So, it is quite important to increase not only the image quality and the sense of speed but also the fun. By using our proposed technique, we can well-achieve this purpose. Therefore, this technique is effective for expanding field-of-view on large screens.

6. Conclusions

In this paper, we proposed the technique to reconstruct pseudo wide-angle images by using the approximative depth model in real-time. We implemented the technique on the multi-projector display system, and evaluated its characteristics. Through these experiments, we could effectively generate pseudo peripheral images with the approximative depth model in real-time. The expanded images could achieve high immersion and fun compared with other previous methods.

As our future works, we have a plan to use different shaped depth model to achieve fast and accurate approximation. And also, we will try to apply this technique to many kinds of contents.

References

[CNSD93] CRUZ-NEIRA C., SANDIN D. J., DEFANTI T. A.: Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. In *Proc. SIGGRAPH 1993* (1993), pp. 135–142. 1

[EMSN07] EMOTO M., MASAOKA K., SUGAWARA M., NOJIRI Y.: Evaluation of the sense of presence in viewing wide-field images. *The Technical Report of the Institute of Electronics Information and Communication Engineers of Japan (MVE2006-75) 106*, 496 (2007), 25–30. 5

[HEB*01] HUMPHREYS G., ELDRIDGE M., BUCK I., STOLL G., EVERETT M., HANRAHAN P.: Wiregl: A scalable graphics system for clusters. In *Proc. SIGGRAPH 2001* (2001), pp. 129–140. 2

[HEB*02] HUMPHREYS G., ELDRIDGE M., BUCK I., STOLL G., EVERETT M., HANRAHAN P.: Chromium: A Stream Processing Framework for Interactive Rendering on Clusters. In *Proc. SIGGRAPH 2002* (2002). 2

[HHS04] HASHIMOTO N., HASEGAWA S., SATO M.: A Development of Multi-Projection Display "D-vision". *The Journal of The institute of image information and television engineers 58*, 3 (2004), 409–417. 4

[HIS06] HASHIMOTO N., ISHIDA Y., SATO M.: A Self-Distributing Software Environment for Immersive Multiprojector Displays. *The Transactions the Institute of Electronics Information and Communication Engineers of Japan J89-D*, 2 (2006), 362–370. 2

[ISI*06] IKETANI A., SATO T., IKEDA S., KANBARA M., NAKAJIMA N., YOKOYA N.: Video mosaicing for curved documents based on structure from motion. In *Proc. 18th IAPR Int. Conf. on Pattern Recognition* (2006), vol. 4, pp. 391–396. 2

[JBC] JapanBroadcastingCorporation, http://www.nhk.or.jp/. 1

[MS] MicrosoftCorporation, http://www.xbox.com/. 1

[SCE] SonyComputerEntertainmentInc, http://www.playstation.com/. 1

[Sze94] SZELISKI R.: Image Mosaicing for Tele-reality Applications. In *Proc.IEEE Workshop on Applications of Computer Vision* (1994), vol. 1, pp. 230–236. 2