ICAT'98

Techniques for Modeling and Rendering a Video Image-Based Virtual Environment for Driving Simulation System

ZHANG Xiaohui ZHANG Xiaohua Masayuki NAKAJIMA

Graduate School of Information Science & Engineering, Tokyo Institute of Technology
2-12-1 O-okayama, Meguro-ku, Tokyo, 152-8552 Japan
{zhang,zhxh,nakajima}@cs.titech.ac.jp

Abstract

The virtual environment used in driving simulation systems is large and usually contains outdoor scene. Therefore, it is very difficult to model. In addition, it is also hard to render the outdoor scene realistically using a traditional CG rendering pipeline. Recently more and more researchers have begun begin to use Image-Based Rendering(IBR) methods in order to deal with large scenes and improve realism. In this paper, we propose some techniques which use video images to model and render scenes in a driving simulation system. These techniques require the use of forward moving image sequences, which are very easy to obtain. There are three main points: (1) We propose a mosaicing technique for forward moving images (Forward Moving Image Mosaics, or FMIM for short); (2) The virtual views generated by FMIM have different resolutions; (3) The visibility of the view is changeable by adding fog effects. In this paper, we describe our algorithm and the techniques we use in our video image based driving simulation system in detail and show some experiment results.

Key words: Image-based Rendering, Image Mosaics, Virtual View Synthesis, Fog Effects, Driving Simulation

1 Introduction

Virtual driving environments represent a challenging test for virtual reality technology. In a driving simulation, scene database creation and terrain modeling are intimidating problems. It is also very difficult to generate realistic images by 3D model-based rendering, and the speed of 3D model based rendering depends on the complexity of the scene. Moreover for outdoor scenes, it is also difficult to simulate natural light.

Recently, there are emerging and competing methods for creating virtual views called image-based rendering. In contrast with 3D model-based rendering, image-based rendering techniques rely primarily on original images to produce new virtual views.

There are a great deal of background research in image-based rendering. A survey of these techniques can be found in [1]. The representative techniques are image mosaics[2][3], interpolation from reference images[4], light field rendering[5][6], and geometrically valid pixel reprojection [7][8][9][10]. Image mosaics techniques usually generate a panorama image from images taken while rotating the camera around its center of projection or while translating the camera position. Techniques which interpolate from reference images cannot always ensure the generated image is physically valid. Levoy and Hanrahan introduced the light field rendering in it is not necessary to extract model information, such as depth values or image correspondences. But the primary disadvantage of this method is the requirement of a large set of image samples. For geometrically-valid pixel reprojection, as we know, no work has been done using the forward moving images.

Until now, only a few researchers have worked on the illumination problem[11][12] which is one of the most difficult problem of IBR. The calculation of both methods is very time consuming, so that are not suitable for a driving simulation system. Both also need some 3D geometry model of the scene.

Video image is beginning to play an important role as an information source in computer graphics. Particularly, forward moving video images maybe the easiest information to obtain because they need only one video camera, and the motion of the camera is just forward moving. That is the main reason why we study on this type of original images instead of using images taken by multiple cameras from different directions. By considering the distinguishing features of a driving simulation system, we found that a forward moving image sequence is suitable. Based on this background, we propose a new technique which we call Forward Moving Image Mosaics (FMIM) and use it in our driving simulation system.

Our whole system consists of two parts: one is off-line processing, the other is on-line processing. The off-line processing constructs the virtual environment from a forward moving image sequence, while the on-line processing is responsible for producing the virtual views by using FMIM and adding fog effects in real-time.

We begin this paper with the review of related work in Section 2. Then we describe the features of forward moving image sequences and our driving simulation system in Section 3. The off-line and on-line processing are discussed in Section 4 and Section 5 respectively. We show our experiment results in Section 6. Finally, conclusions and future work are given in Section 7.

2 Related Work

Richard Szeliski and Sing Bing Kang have done much work on image mosaics. In [3], they presented the concept of video mosaics. In their techniques, they solve a least-squares estimation problem in the unknown structure and motion parameters, and give algorithms for both constructing planar and panoramic mosaics and for projective depth recovery. But their method is only suitable for two specific imaging situations. One is when the images cover a portion of a planar scene. The other is when the camera rotates around an axis through its focal point. In fact for an outdoor scene the first case is hard to satisfy, and the second case limits the camera to a fixed viewpoint. In addition they did not mention forward moving image sequences.

In [13], they extended the algorithm of TIP (Tour Into Pictures)[14] to forward moving image sequences in order to improve the resolution. In their method, 3D pseudo-model is constructed by combining a series of pseudo-boxes. Since TIP requires the decomposition of the image into five parts manually, the work of 3D pseudo-model construction is still time consuming. In addition, because it is 3D pseudo-model, the virtual view is not guaranteed to be physically correct. Moreover this method is not very suitable for outdoor scenes.

Also [15] uses video sequence to construct 3D natural scenes. In this method, a spatio-temporal texture is employed to construct the scene as a set of panoramic depth layers, each of which consists of an intensity map and a depth map. But this method is also not suitable for forward moving image sequences.

In [16] also uses video images for sports simulator, Peloton. The authors incorporate video displays into virtual environments. Our techniques is similar to theirs, but we do not use any 3D virtual environment at all. Therefore the virtual views we generate are based only on images and have no artificial marks.

None of these related works [13][14][15][16] dealt with how to change the illumination. In this paper we employ the forward moving image sequence to model the natural scene and use the FMIM technique to produce virtual view in driving simulation system. During the rendering process, by using our fog generation algorithm, we can generate virtual images in different moisture situations. In a word, our video images based driving simulation system can not only generate realistic virtual view in real time, but also can change the visibility in order to train the driver in different conditions.

In the following section, we discuss these techniques and our proposed solutions.

3 Features of Forward Moving Image Sequences and a Driving Simulation System

In this section, we first describe the features of forward moving image sequences and how to acquire original forward moving images. Then we briefly discuss the special features of a driving simulation system.

3.1 Features of Forward Moving Image Sequences

A forward moving image sequence is an array of images taken by moving a camera forward along its optical axis, shown in Figure 1.

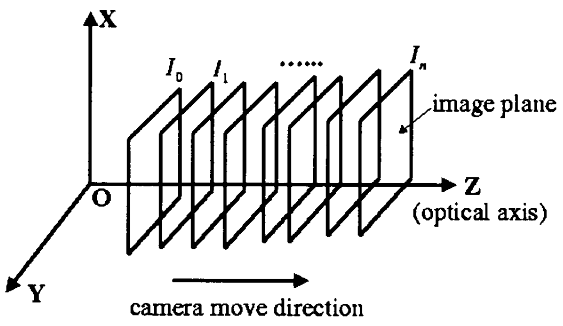


Fig.1 Forward moving image sequence

Let I be the symbol for an image, and S represent the scene which is in the camera frustum. Suppose that I_0 and I_1 are two consecutive images, and S_0 and S_1 are their corresponding scenes. Without occlusion, it is obviously that $S_1 \subset S_0$, as Figure 2 shows. Even though $S_1 \subset S_0$, it does not mean that I_0 has all the information contained in I_1 because image I_1 contains the details of far objects which image I_0 does not. That shows the field of view of one image is limited, but the wider view of information is included in the forward moving image sequence. This feature is what we want to use.

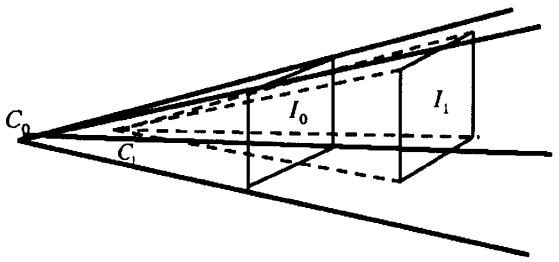


Fig.2 Forward moving image frustum

We obtained a forward moving image sequence by mounting a video camera to a hand-cart, and kept it moving straight and at an uniform speed, as shown in Figure 3. By assuming the road is flat, we regard the image sequence taken this way as forward moving images.

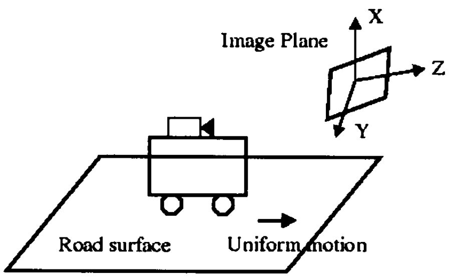


Fig.3 Photographing process

3.2 Features of a Driving Simulation System

By analyzing the driving simulation system, we found that there are two special features in driving simulation system:

1. the viewpoint does not change violently
2. the changes of the viewpoint are almost horizontal

Because of these features, we say that the forward moving image mosaics technique can be applied to a driving simulation system effectively.

4 Off-line Processing

The off-line processing is the process of construction virtual environment. In this part, the camera's position and the depth map of each frame should be recovered. The output of the off-line processing is the virtual environment which is composed of original images, depth maps and camera positions.

4.1 Camera Position

We have to extract the camera position (C_x, C_y, C_z) and focal length f for each frame. While taking the video, we keep the focal length fixed, so we only need to calibrate the camera position (C_x, C_y, C_z) for each frame.

We set the camera position of the first frame as the original point O, and the optical axis as the Z axis. Let L be the distance which video camera moves from the start point to the end point and N be the frame number. Since the motion of the video camera is uniform, the distance $\triangle C_z$ between two consecutive camera position along the motion direction is

$$\triangle C_z = \frac{L}{N}$$

Obviously, the camera position (C_x^i, C_y^i, C_z^i) of frame i can be calculated as follows:

$$\begin{aligned} C_x^i &= 0 \\ C_y^i &= 0 \\ C_z^i &= i \times \triangle C_z \end{aligned} \qquad (1)$$

4.2 Depth Recovery

The coordinates corresponding between a 3D point $P_i = (X_i, Y_i, Z_i)$, and the its pixel (x_{ij}, y_{ij}) in the jth image can be written as

$$\begin{aligned} x_{ij} &= f \cdot \frac{X_i}{Z_i - j\triangle C_z} \\ y_{ij} &= f \cdot \frac{Y_i}{Z_i - j\triangle C_z} \end{aligned} \tag{2}$$

where j is the frame number, i is the pixel number. Suppose pixel (x_{pq}, y_{pq}) in the qth frame corresponds to pixel pixel(x_{ij}, y_{ij}), i.e. they correspond to the same point in 3D space. From (2), we can deduce the following:

$$\begin{aligned} x_{pq} &= \frac{Z_i - j\triangle C_z}{Z_i - q\triangle C_z} \cdot x_{ij} \\ y_{pq} &= \frac{Z_i - j\triangle C_z}{Z_i - q\triangle C_z} \cdot y_{ij} \end{aligned} \tag{3}$$

The above formulation extends naturally to multi-frame depth recovery by simply minimizing the summed intensity error

$$E = \sum_{j \neq 0} \sum_{i} [I(x_{ij}, y_{ij}) - I(x_{i0}, y_{i0})]^2 = \sum_{j \neq 0} \sum_{i} e_{ij}^2 \tag{4}$$

Using the Leverberg-Marquardt algorithm[17], Z_i for each pixel can be obtained. In our system, we use three successive frames to recover the depth map of the middle frame.

5 On-line Processing

The on-line processing produces the virtual view corresponding to the user's viewpoint and can add fog effects according to user's setting. There are three situations concerning the user's viewpoint.

1. Both position and direction are the same as one of the original images;
2. The position is the same as one of the images, but with a different direction;
3. Neither the position nor the direction is the same as any of the original images.

The first situation is very simple because we just need to output the corresponding original image. For the second and third situations, the FMIM technique is employed. In this section, we focus on our FMIM and introduce our fog effects algorithm.

5.1 Forward Moving Image Mosaics

Taking advantage of the conclusions drawn in Section 3.1 that a larger field of view is contained in the forward moving image sequence, a virtual view with larger field of view than the original image and with different resolution can be generated. We call this technique *Forward Moving Image Mosaics* (FMIM). This technique has two steps: depth image reprojection and image overlaying.

5.1.1 Depth Image Reprojection

After we know the parameter of the camera and the depth (Z) values at each pixels, it is straight forward to reproject the image onto a new image plane. If 4×4 matrix C_1 describes the first camera, and C_2 is the desired view, we have

$$\begin{bmatrix} x_{i1}w_{i1} \\ y_{i1}w_{i1} \\ z_{i1}w_{i1} \\ w_{i1} \end{bmatrix} = C_1 \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}, \quad \begin{bmatrix} x_{i2}w_{i2} \\ y_{i2}w_{i2} \\ z_{i2}w_{i2} \\ w_{i2} \end{bmatrix} = C_2 \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}$$

and so

$$\begin{bmatrix} x_{i2}r \\ y_{i2}r \\ z_{i2}r \\ r \end{bmatrix} = C_2 C_1^{-1} \begin{bmatrix} x_{i1} \\ y_{i1} \\ z_{i1} \\ 1 \end{bmatrix} \tag{5}$$

Here we only consider the case where new view has the same frustum but a different position. Specifically, the new view position is just a translation in Z direction. Thus, the equation (5) can be simplified as follows

$$\begin{aligned} x_{ki}^{j} &= \frac{Z_k}{Z_k - \triangle C_z \cdot (j - i)} x_{ki} \\ y_{ki}^{j} &= \frac{Z_k}{Z_k - \triangle C_z \cdot (j - i)} y_{ki} \end{aligned}$$

Here (x_{ki}, y_{ki}) is kth pixel in the image I_i with the camera position $(0, 0, i \cdot \triangle C_z)$, and Z_k is the depth value of pixel (x_{ki}, y_{ki}). By reprojection pixel (x_{ki}, y_{ki}) to a new view with camera position $(0, 0, j \cdot \triangle C_z)$, we obtain the pixel (x_{ki}^{j}, y_{ki}^{j}). Thus a virtual image I_i^j is generated, as shown in Figure 4. The size of virtual image I_i^j becomes larger and consequently the resolution decreases.

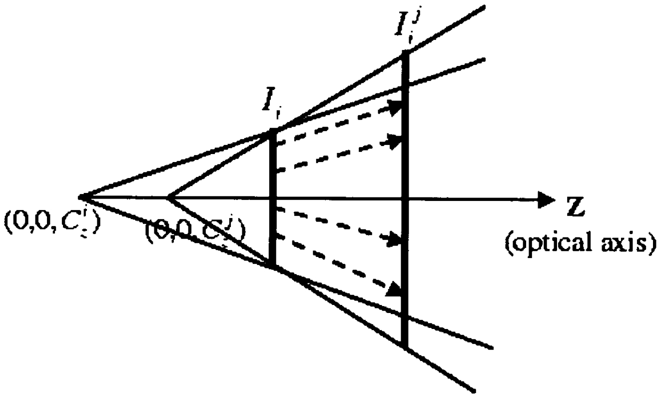


Fig. 4 Forward depth reprojection

Since Z_k is dependent on (x_{ki}, y_{ki}), the inverse mapping cannot be determined explicitly. We use a forward mapping technique and employ super-sampling[18] to alleviate the effect of noisy estimation in the depth map. There are still some holes after the forward mapping because there is no corresponding pixel in the source image. We just interpolate the nearby pixels to fill the holes.

5.1.2 Overlaying Images

Even though image I_i^j has larger field of view than the original image, its resolution is low, especially for distant objects. There are also some objects which can be seen from $(0, 0, j \cdot \triangle C_z)$, but can not be seen from $(0, 0, i \cdot \triangle C_z)$. We solve this problem by overlaying the source image I_j whose camera position is $(0, 0, j \cdot \triangle C_z)$ to I_i^j and generate a new image $I_i^{j\prime}$. As figure 5 shows, after overlaying we get a virtual image which not only has larger field of view but also has multiple resolution. In the center the resolution is high, while in the boarder, the resolution is low. This is just what we want, because in a driving simulation system human eyes usually see the center more clearly than the border. The image $I_i^{j\prime}$ is called a *forward moving mosaic image.*

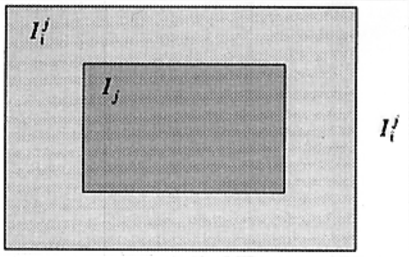


Fig. 5 Overlay I_j to I_i^j. The resolution in the center is higher than that of in the border.

5.2 Application of FMIM to a Driving Simulation System

As we have said, a driving simulation system has the features that the viewpoint does not change too much and the movements are almost always horizontal. It is because of these features that we can apply the FMIM technique.

Suppose the user's position is $(C_x^j, 0, j \cdot \triangle C_z)$, and at this position, he changes his viewpoint by rotating 15 degree horizontally as Figure 6 shows. It is obvious that only rotating the source image I_j only cannot satisfy the new view because it needs more information which is not contained in the source image I_j. In Section 3, we illustrate that the wider field of view is contained in the forward moving image sequence. Therefore we can make use of the image I_i, the proceeding image to image I_j. Now the problem is how to select image I_i.

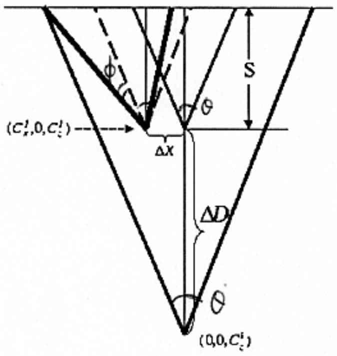


Fig. 6 Method for selecting image I_i

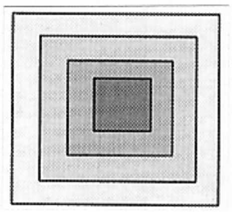


Fig.7 Multi-resolution FMIM

Let θ be the view angle of the original image (it can be calculated from focal length f and the size of the image) and ϕ be the rotation angle. Suppose we are only concerned scenes which are S far away. From Figure 6, it is easy to know that the image which is taken on $\triangle D$ before image I_j contains the information the new view needs, and $\triangle D$ can be calculated

$$\triangle D = \frac{S \cdot tg(\frac{\theta}{2} + \phi) + \triangle X}{tg(\frac{\theta}{2})} - S \tag{6}$$

where $\triangle X = C_X^j$.

After $\triangle D$ is calculated, image I_i can be selected using equation (1). Using the FMIM technique to reproject image I_i to the position $(0, 0, j \cdot \triangle C_z)$, we get a larger field of view image. Next we overlay image I_j to this larger image and obtain image $I_i^{j\prime}$. When we get image $I_i^{j\prime}$, by using the translate and rotate transform, the desired image can be obtained.

If the rotation angle ϕ or C_X^j is very large, from equation (6), we know that $\triangle D$ is also very large. In this case, the final result is not good because of the low resolution. If we perform the FMIM more than once by using small $\triangle D$ instead of executing it only once with a very large $\triangle D$, a high quality image can still be generated.

From the above, we can draw the conclusion that if the position of user's viewpoint is the same as one of the original image, but with a different direction, the virtual image generated by the FMIM is physically valid. Since the viewpoint is almost always moving forward in driving simulation system, FMIM is very suitable.

5.3 Adding Fog Effects

In a driving simulation system, it is necessary to simulate various weather conditions, such as fog, rain, snow etc. Controllable illumination is still one of the difficult problem for IBR because geometrical and lighting information are not sufficient or are hard to extract. In this paper, we propose an approach to give fog effects.

Fogging can create atmospheric effects and give depth cuing which are very important in driving simulation system. Fogging is a non-linear function of distance[20]. With fog, objects far from the viewpoint seem to lose their color and tone. The reduction rate of color and tone depends mainly on density and distance, by the following equation:

$$r = e^{-(density \cdot z)} \tag{7}$$

Here r is a symbol for the reduction rate of color and tone, *density* specifies the fog's density (here we use the constant fog model, i.e. has a constant density everywhere), and z is the distant from the viewpoint.When adding fog effects, the color c_{ij} of pixel i in j frame becomes to $c_{ij'}$

$$c_{ij'} = rc_{ij} + (1-r)C_{fog} \tag{8}$$

where C_{fog} is the color of fog.

Since we have a depth map for each frame, we can use it directly to add fog effects. By adjusting *density* and C_{fog}, we can simulate different fog conditions or even different times of day to some extent. For example, by setting C_{fog} to dark gray, we can simulate the scene in dusk with fog.

6 Experimental Results

We collected a series of images using a video camera on a leveled tripod in the road very near to our university. While taking the video, we kept the hand-cart moving straight and at an uniform speed and maintained a constant focal length. The frames were then digitized at a rate of approximately 5 frames per second to a resolution of 320 by 240 pixels. An example of three sequential frames are shown in Figure 8.

Then we recovered the depth for each frame. By using the FMIM technique, we obtained an image with a larger field of view and multiple resolutions (Figure 9(a)). Then employing the rotate transformation, we formed the virtual view, as Figure 9(b) shows. Figure 9(c) gives a view image with fog effects.

7 Conclusion and Future Work

FMIM is a technique which is very efficient for forward moving image sequences. The generated virtual views generated have multiple resolutions, which is the main point of this technique. Moreover the quality of depth map does not effect the quality of the final virtual image as in other techniques because the source image overlays the reprojected image. In fact, it is not necessary to perform the reprojection on the whole image. Operating on the image border is enough. Thus the calculation can be reduced greatly. For weather conditions, although we only deal with the fog, it is a promising beginning work.

Overall, our video image-based driving simulation system has mainly three novel contributions: (1) It makes use of the forward moving image sequence which is seldom studied or used as source data. (2)We proposed the FMIM technique. This technique is very suitable for a driving simulation system. (3)Fog effects are added to the virtual view almost without increasing the amount of computation.

Occlusion is not considered in this paper. In addition the forward moving image mosaics technique requires that the source images must be forward moving images. As we know that occlusion problem always exits and the road also cannot always be straight. Therefore we have to solve the problem of occlusion and how to apply this approach to non-forward moving images in order to make this technique widely applicable. Also, in this research, we only dealt with the fog effects. We will make an effort to generate other weather conditions in the future.

This paper was supported by TAO project *The Study of Composition, Editing and Retrieval of Image Contents in the Next Image Generation System*

References

[1] S.B.Kang, "A Survey of Image-based Rendering Techniques," Technical Report CRL 97/4, Cam-

bridge Research Lab., Digital Equipment Corp., August 1997.

[2] Shum Heung-Yeung, Richard Szeliski, "Panoramic Image Mosaics," Technical Report. MSR-TR-97-23. http://www.research.microsoft.com/.

[3] R.Szeliski,"Video Mosaics for Virtual Environment," IEEE Computer Graphics and Applications, March 1996. pp. 22-30.

[4] Chen,S.E., Willians,L., "View Interpolation for Image Synthesis," Proc. SIGGRAPH '93 (Anaheim, California, August 1-6,1993). In Computer Graphics Proceedings, Annual Conference Series, 1993. ACM SIGGRAPH, pp. 279-288.

[5] Marc Levoy, Pat Hanraham, "Light Field Rendering," SIGGRAPH'96, Aug. 1996 pp. 31-41.

[6] S.J. Gortler, R. Gzeszczuk, R. Szuliski, and M.F.Cohen, "The Lumigraph," SIGGRAPH'96, Aug. 1996, pp. 43-54.

[7] O.D. Faugeras, "What Can Be Seen in Three Dimensions with an Uncalibrated Stereo Rig?" in Second European Conference on Computer Vision(ECCV'92), Springer-Verlag, Santa Margherita Liguere, Italy, May 1992, pp. 563-578.

[8] N.L. Chang, A. Zakhor, "View Generation for three-Dimensional Scenes from Video Sequences," IEEE Transactions on Image Processing, 6(4) April 1997, pp.584-598.

[9] S. Laveau and O. Faugeras, "3-D Scene Representation as a Collection of Images and Fundamental Matrices," Technical Report 2205, INRIA-Sophia Antipolis, Feb. 1994.

[10] A.Shashua, "Algebraic Functions for Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(8):779-789, 1995.

[11] Paul Debevec, "Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-based Graphics with Global Illumination and High Dynamic Range Photography," SIGGRAPH'98, pp. 189-198.

[12] Yizhou Yu, Jitendra Malik, "Recovering Photometric Properties of Architectural Scenes From Photographs," SIGGRAPH'98, pp. 207-217.

[13] Zhang Xiaohua, "Construction of Virtual Environment from Video Data with Forward Motion," The 1st International Conference on Advanced Multimedia Content Processing(AMCP'98), Oosaka, Nov.,1998,pp. 306-318.

[14] Y.Horry, K. Anjyo, K.Arai,"Tour Into the Picture: Using a Spidery Mesh Interface to Make Animation from a Single Image," Proc. SIGGRAPH '97 (Los Angels, California, August 3-8,1997). In Computer Graphics Proceedings, Annual Conference Series, 1997. ACM SIGGRAPH, pp. 225-232.

[15] Zhigang Zhu, Guangyou Xu, Xueyin Lin, "constructing 3D Natural Scene from Video Sequences with Vibrated Motion," IRAIS'98, pp. 105-112.

[16] Gianpaolo U. Carraro, John T.Edmark, J. Robert Ensor, "Techniques for Handling Video in Virtual Environments," SIGGRAPH'98,pp. 353-360.

[17] W.H.Press, B.P.Flannery, S.A. Teukolsky, and W.T.Vetterling. "Numerical Recipes in C: The Art of Scientific Computing," Cambridge University Press, Cambridge, England, second edition, 1992.

[18] G. Wolberg. "Digital Image Warping," IEEE Computer Society Press, Los Alamitos, California, 1990.

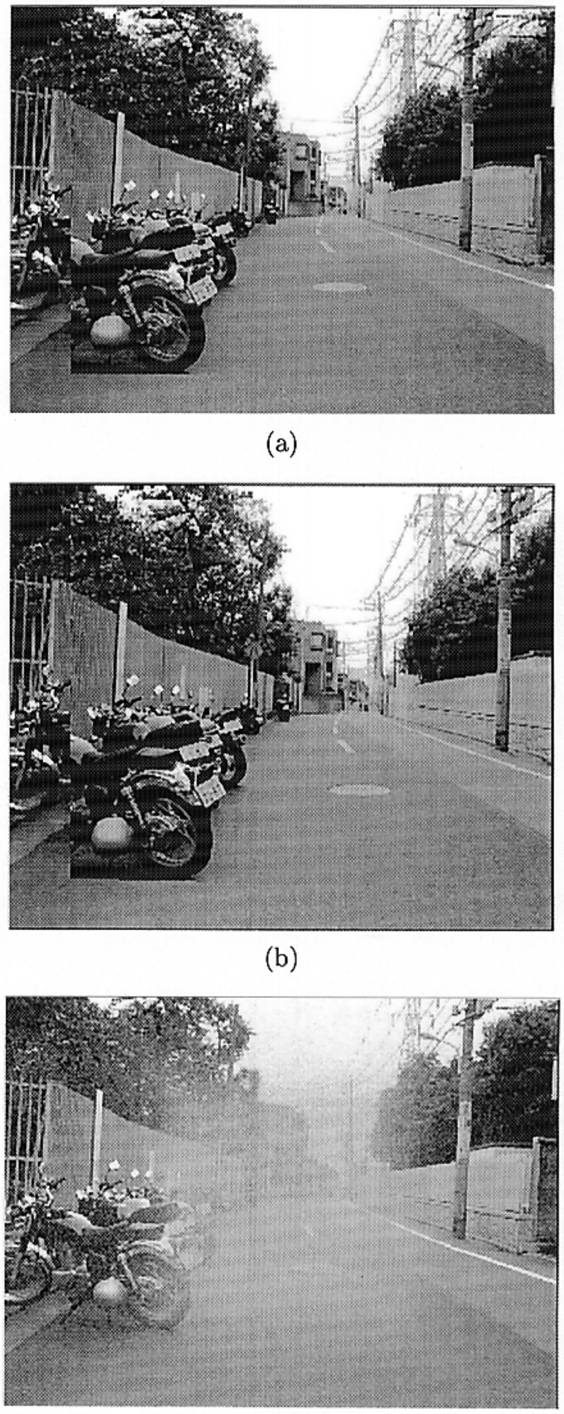
[19] R. Victor Klassen, "Modeling the Effect of the Atmospher on Light," ACM Transactions on Graphics, Vol. 6, No3, July 1987, pp. 215-237.

frame 1

frame 2

frame3

Fig. 8 Three frames of a forward moving image sequence



(a)

(b)

(c)

Fig. 9 (a)Results of FMIM (b)Virtual view image (c)Fog effects