

GENERATION OF MULTIVIEWPOINT VIDEO FROM STEREOSCOPIC VIDEO

Ho-Chao Huang¹, Ching-Che Kao², and Yi-Ping Hung¹

¹Institute of Information Science, Academia Sinica, Taipei, Taiwan

²Department of Computer Science and Information Engineering
National Taiwan University, Taipei, Taiwan

Abstract

This paper proposes a new technique for generating multiviewpoint video from two-view stereo video sequence. The two-view stereo video can be easily obtained by using inexpensive two-view stereo video capture devices. For each image pair in the two-view stereoscopic video signals, our system first estimates the corresponding points of each pixel based on epipolar constraint. A smoothing algorithm is used to smooth the estimation result, and then generate the disparity maps for each image pair. Then the proposed system can generate multiple perspective stereo video by interpolating or extrapolating the original views based on the generated disparity maps. Compared to the traditional method of capturing the multiple perspective video directly, Our method can significantly reduce both the difficulties of video capturing and processing as well as the amount of video data.

1 Introduction

For many immersive entertainment[1] and virtual reality systems[2, 3, 4, 5, 6], providing stereo views is an important feature to let users feel as if they are participating, especially if the systems can provide multiple perspective stereo views. Many new stereo monitor systems have the capability to simultaneously project video signals of multiple, e.g., six or more, perspectives. How to obtain and how to manage multiple perspective video is an important issue for those systems. However, it is not easy to capture multiple, e.g. six, perspective video because multiple perspective video capture device is expensive and unpopular. Even if one can capture multiple perspective video that are synchronized correctly using special hardware, it is still not easy to store and manipulate those video sequences due to the large amount of image data.

In this paper, we propose a technique for generating multiple perspective video from a two-view stereo video sequence. Compared with multiple perspective video capture device, stereo video camera is much more pop-

ular and less expensive. In the proposed mechanism, we first obtain the initial two views using a stereo video camera. The next step is to estimate the corresponding points of each pixel in the stereo video automatically or semi-automatically, and then generate two disparity maps for each image pair in the stereo video. The generation of disparity maps is a time consuming procedure, but it can be done by off-line processing. By applying the proposed estimation algorithm, the disparity maps for most stereo videos can be generated fully automatically. From the original stereo video and the generated disparity maps, our system can generate multiple perspective stereo views on-line by interpolating or extrapolating the original stereo video. By using the proposed technique for multiple views generation, only the original stereoscopic video signals and the generated sequences of disparity map are needed for generating a sequence of stereo video which allows user to change viewpoint within a certain range around the original two views. Both the difficulty of video processing and the amount of video data can be reduced significantly.

2 Review

This section briefly reviews the multiple perspective video display system and some morphing techniques commonly used for view interpolation.

Multiple perspective video display system

Several kinds of stereo display systems are commonly used in many multimedia applications. For example, the polarized stereo video projecting system is widely used in theater to show stereo movies. The audiences must wear a polariscope stereo glasses to enjoy the stereo movie. In computer display systems, LCD shutter glasses is commonly used to see stereo objects on computer monitor. And the blue-red stereo glasses is widely used to see stereo objects on computer monitor, slide shows and printed articles. All the commonly used stereo display mechanisms can only display two perspective views at

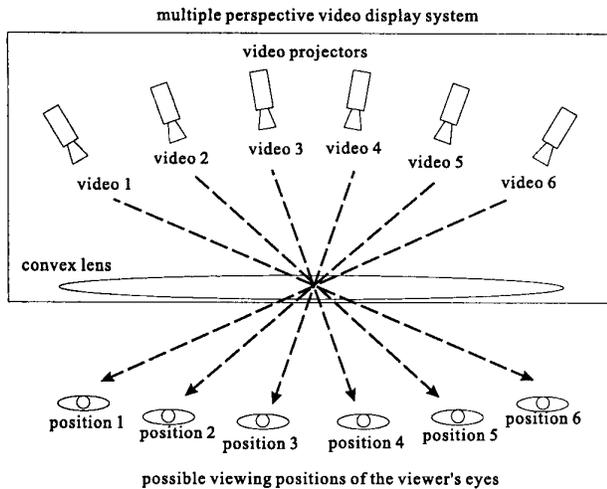


Figure 1: The multiple perspective video display system.

the same time, and the users need to wear some kind of stereo glasses to see the stereo output.

The multiple perspective video display system is an alternative method to show stereo videos. Figure 1 is the brief diagram of an example multiple perspective video display system containing six video projectors. In figure 1, The six video projectors numbered from one to six project different perspective videos on the convex lens. Thus the images of those videos are located at the front of the display system in reverse order. There are six possible position to see the projected videos. When the user's eye is at position 1, the user can see the video projected by video projector 6, for example, and if the user's eye is at position 2, he can see the video 5, and so on. Most of the multiple perspective video display systems are well designed such that the user's left and right eyes can be at the adjacent viewing positions at the same time. For example, the user can put its left and right eyes at viewing position 3 and 4, respectively, to see the videos 4 and 3 simultaneously to experience the stereo movie. The user can move his head to left or right. For example, when the user's head moves to left, his left and right eyes move to viewing position 2 and 3, respectively, and the observed videos are switched to video 4 and 5, simultaneously, to see another stereo view. That is, the example display system shown in figure 1 can provide five stereo perspective views simultaneously, and the users need not to wear any kind of stereo glasses to see the stereo video.

The multiple perspective video display system is a better device for immersive systems because of the advantages described above. However, it is not easy to obtain the multiple perspective video, and it is not easy to manipulate the video data. The goal of this paper is to automatically generate multiple perspective views from two views by using morphing technique, and it is useful for the multiple perspective video display system.

Review of morphing techniques

Morphing technique contains two transition functions for each corresponding pair in the source object and the target object. One is the forward transition function, which defines how the corresponding point moves from the source object to the target object. The other is the backward transition function, which defines how the corresponding point moves from target object back to the source object. The forward transition function generates an intermediate object from the source object, and the backward transition function generates an intermediate object from the target object. In both functions, there are a variable s varying from 0 to 1, which determines that the intermediate object is how far away from the source object. When s is 0, the position of intermediate object is the same as that of the source object. When s is equal to 1, the position of intermediate object is the same as that of the target object. The corresponding pairs of the forward transition function need not to be the same as those of the backward transition function. Morphing technique also contains an interpolation function, which determines how the intermediate objects of the forward and backward transition functions are merged into resulted transition objects.

The image morphing[7], in which the "object" is equal to an image, is able to be applied to the image-based rendering systems. The image morphing contains a set of functions which morph each pixel from the source image to the target image. However, traditional image morphing function is usually a linear transition function of 2D image positions. If we simply apply the image morphing to an image-based virtual reality system, the generated transition views may not be the same views the viewers can see when they walking from the source viewpoint to the destination viewpoint in the real world.

The view morphing[8] technique was proposed to generate realistic transition views for lateral transition function. View morphing contains three steps to complete its work described as follows.

The first step of view morphing is prewarping. The fundamental matrix[9, 10, 11] for the positions relation of the camera when taking the source and target images is estimated from the corresponding features in the source and target images. Then the source and target images are warped onto parallel image planes, and both warped image planes are parallel to the transition baseline.

The second step of view morphing is to apply traditional image morphing with linear transition function on the warped images. After the transition images are generated by applying the linear image morphing, those images are then back projected to the image planes, such that the resulted images are almost the same as the images taken by camera moving from the original source position and viewing direction to the original target position and viewing direction.

The disparity morphing has been proposed in [12] to

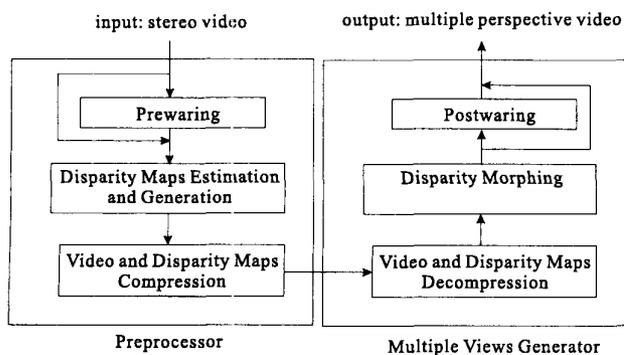


Figure 2: The block diagram of the multiple perspective video generation system.

generate realistic transition views from two images for general view transition in image-based rendering systems. The disparity morphing depends on the estimated disparity map for each image pair based on the epipolar geometry. The epipolar constraints can not only speed up the searching of correspondences for disparity estimation, but also improve the accuracy of the estimated disparity maps. Then the forward and backward morphing functions are defined based on the estimated disparity maps. Notice that the forward and backward morphing functions are usually not linear and not similar to the traditional image morphing functions. Finally, the interpolation function is defined to deal with the object occlusion problems correctly. The resulted transition views are almost the same as what viewers can see in the real world.

3 Multiple Perspective Video Generation

This section describes the proposed multiple perspective video generation system. The video generation algorithm used in this system is based on the disparity morphing technique proposed in [12]. The block diagram of the video generation system is shown in figure 2.

As shown in figure 2, the original stereo video captured by stereo video camera is the input data of the views generation system. The preprocessor of the multiple views generation system prewarps the stereo video such that the video planes are parallel to each other, then estimates the correspondences of each pixel in the stereo video, generates the disparity maps, and then compresses the original stereo video as well as the generated disparity maps. The procedures of the preprocessor can be done in advance. The multiple views generator of the proposed system reads and decompresses the original videos and disparity maps, then applies the disparity morphing to generate desired views, and then performs the postwarp-

ing procedure to put the output video at desired viewing direction. All the function block of the multiple perspective views generation system will be described in detail in the following subsections from section 3.1 to 3.5.

3.1 Prewarping

The purpose of the prewarping procedure is to make the image planes of original videos be parallel to each other and parallel to the stereo baseline between the optical centers of two cameras[9, 10, 11]. As documented in previous research works[8, 12], once the stereo images pair be parallel to each other and parallel to the stereo baseline, the correspondence estimation can be done by searching on horizontal direction only, and the morphing function becomes linear. Searching on horizontal direction simplifies the estimation procedure and increases the estimation accuracy. The linear morphing function simplifies the views generation procedure and increases the processing speed.

Usually, the original stereo image planes are parallel or almost parallel, and our system has the capability to tolerate the almost parallel cases. Thus the prewarping procedure can be usually bypassed.

3.2 Generation of Disparity Maps

This subsection describes the algorithms for estimating the stereo correspondence and generating the disparity maps. After the original stereo videos are captured by a stereo video camera, an optional pre-warping procedure is applied to warp the stereo video, such that the resulted image planes of two views are almost parallel to each other and parallel to the stereo baseline. Figure 3 shows an example of the stereo image pair, which is generated by 3D Studio for simulation and illustration purposes.

The disparity maps generation algorithm includes two procedures. The first one is the template matching procedure and the second one is the refinement procedure.

Template matching procedure

Since the image planes of the original stereo views are approximately parallel to each other, the corresponding point for each pixel can be searched horizontally based on the epipolar constraint. Due to the search range is constrained within a horizontal area, the correctness of the corresponding point estimation can be increased. In our system, after manually setting the search range, the template matching procedure can automatically estimate the corresponding points. Figures 4 and 5 show some results of the template matching procedure from left image to right image. Figure 4 shows the matching results of low-ambiguity areas, such as the purple box (4a), the background sky (4b), and the textured sphere (4c). Since those areas in figure 4 are distinguishable, the matching results are pretty good. On the other hand,

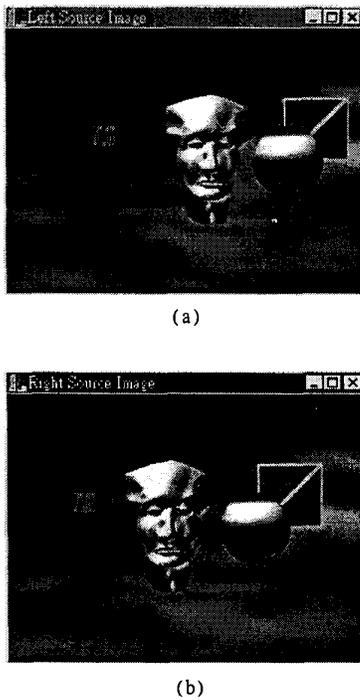


Figure 3: An example of the stereo image pair: (a) left image, (b) right image.

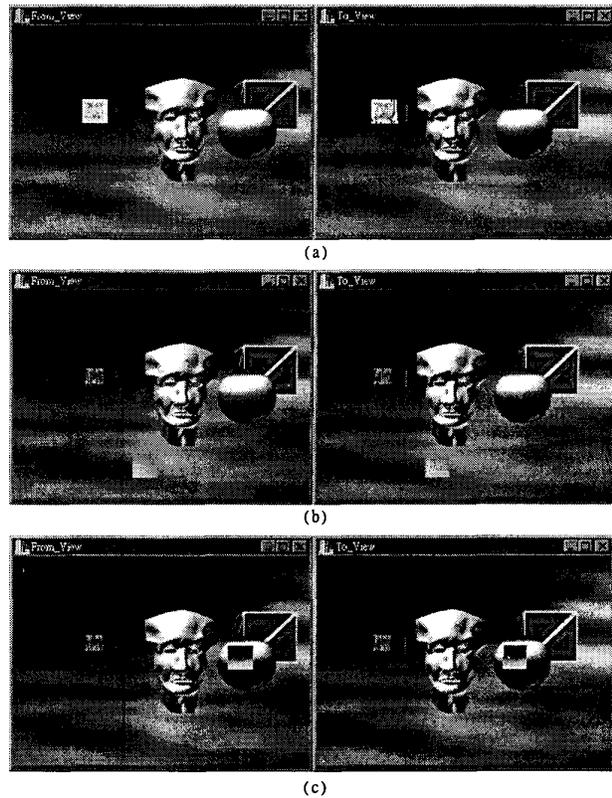


Figure 4: Some results of applying the template matching procedure to low-ambiguity areas.

figure 5 shows the matching results on high-ambiguity areas, such as the face with hollow eyes (5a) and the triangle textured box (5b). The template matching procedure applied on the high-ambiguity areas produces noisy results, and the results should be refined for further uses.

Refinement procedure

Even base-on the epipolar constraints, the result of corresponding point estimation is still noisy and matching ambiguity still can not be fully eliminated. Therefore, a global refinement algorithm is used to remove the matching ambiguity and to smooth out the matching result. The global refinement algorithm is an iteration of two procedures, the median filtering procedure and the smoothing optimization procedure. The median filter uses a cross filter with 50 pixels window size to remove the outliers of the matching results. The smoothing optimization procedure minimizes the smoothness error function, which is the weighted sum of the intensity difference for each corresponding pair, and the matching distance variation of adjacent pixels. The smoothness error function can be used to remove the matching ambiguity and smooth out the matching results, and the function is defined as follows.

$$E = EI + r \times ED, \tag{1}$$

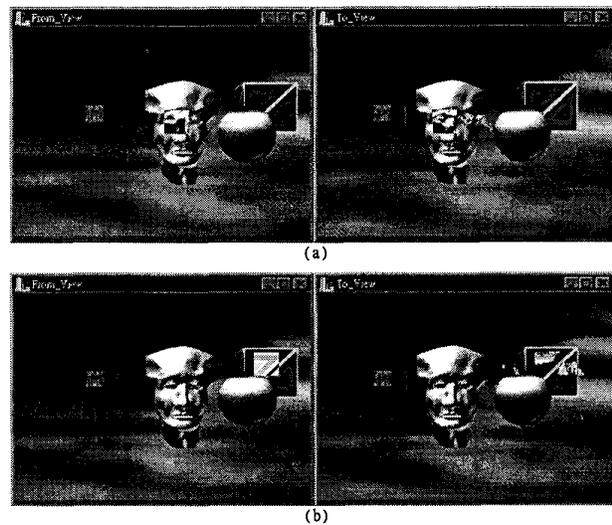


Figure 5: Some results of applying the template matching procedure to high-ambiguity areas.

where EI is the sum of intensity difference of the correspondence defined in equation (2), ED is the matching distance variation defined in equation (3), and r is the refinement factor specified by user and is usually set as 1.

$$EI = \sum_{x,y} \|I(x,y) - I(C(x,y))\|, \quad (2)$$

where $I(x,y)$ is the intensity of pixel at image position (x,y) , and $C(x,y)$ is the corresponding point position of pixel (x,y) .

$$ED = \sum_{x,y} (D(x,y) - D(x-1,y))^2 + D(x,y) - D(x,y-1)^2, \quad (3)$$

where $D(x,y)$ is the distance of corresponding points pair, defined as $D(x,y) = \|(x,y)C(x,y)\|$.

The median filtering procedure and the smoothing optimization procedure are iteratively performed alternately until the resulted smoothness error value E is less than a threshold or the number of iteration times reaches a preset value.

The results of the refinement procedure are shown in figures 6 and 7. Even at the high-ambiguity areas, (7a) and (7b), as well as the object boundaries (7c), the resulted corresponding points are correct.

Generation of disparity maps

After the matching result is refined, the disparity maps can be generated by simple computing the horizontal difference of each corresponding point pair. In our system, the disparity maps are generated as gray images, which can be compressed by using lossless data compression procedure. Figure 8 shows the generated disparity maps for the stereo image pair. The left disparity map shown in figure 8(a) is the disparity map from left image to right image shown in figures 3(a) and 3(b), respectively. On the other hand, the right disparity map shown in figure 8(b) is the disparity map from right image to left image shown in figures 3(b) and 3(a), respectively. The dark value in the disparity map means the position of the corresponding point is at the left hand side of the original pixel position, and the light value in the disparity map means the position of the corresponding point is at the right hand side of the original pixel position. The left and right disparity maps are highly correlated, and the correlation between two disparity maps can be removed by compressing the disparity maps using prediction technique.

3.3 Data Compression

In our system, both the original stereo video and the disparity maps are compressed. The stereo video is compressed using MC/DCT-based video codec[17, 18, 19, 20, 21]. The frame structure is shown in figure 9. The left

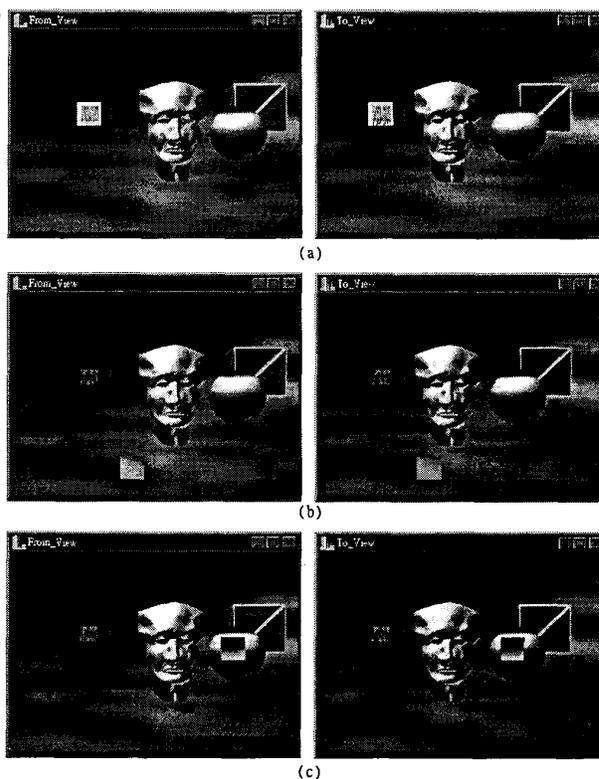


Figure 6: Some results of applying the refinement procedure to low-ambiguity area.

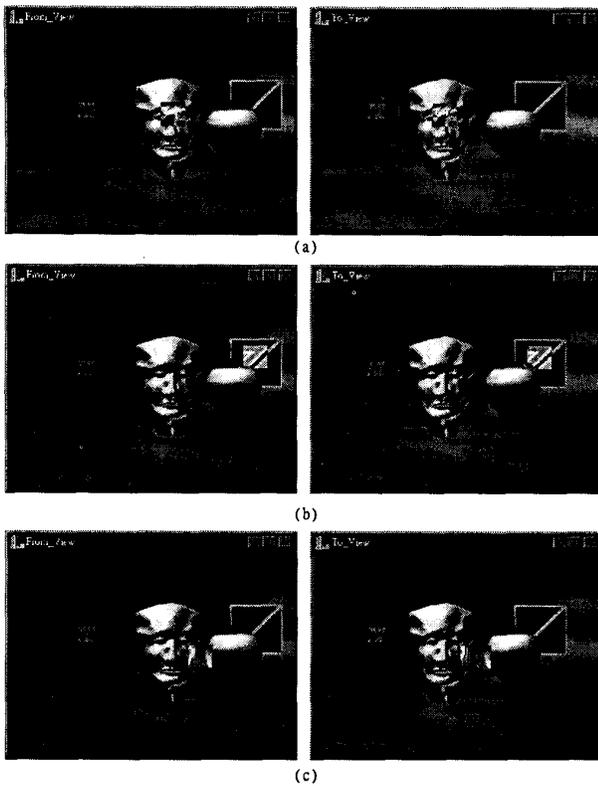


Figure 7: Some results of applying the refinement procedure to high-ambiguity areas.

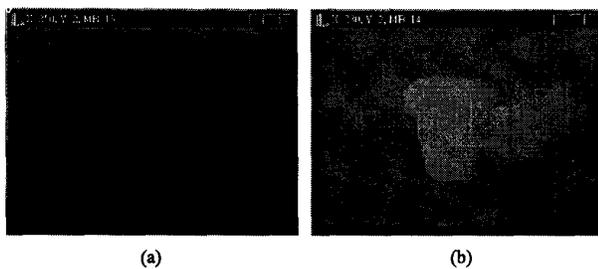


Figure 8: The generated disparity maps for the stereo image pair shown in figure 3: (a) the left disparity map, (b) the right disparity map.

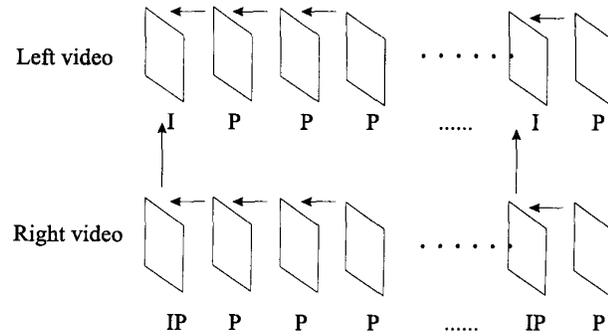


Figure 9: The frame structure of stereo video compression.

video is the reference video. The pictures of the left video are organized as groups of pictures (GOP). The first picture of each GOP is the *I*-picture, which is encoded by DCT transform, linear quantizer, run-length coder, and finally a dynamic arithmetic encoder. The other pictures in each GOP are the *P*-pictures, which are estimated by a motion estimator referencing the previous picture, and the estimation error images are encoded similar to the *I*-picture. The right video is the predicted video. The pictures of the right video are also organized as GOPs. The first picture of each GOP in the right video is the *IP*-picture, which is estimated by referencing the corresponding picture of left video, and the estimation error images are then encoded. The rest pictures of each GOP in the right video are the *P*-pictures, which are encoded as the same as the *P*-pictures of the left video.

The disparity maps are encoded by using lossless technique. The disparity of each corresponding pixel is firstly predicted by the disparity of the previous pixel or the disparity of the corresponding point, and then the prediction error is encoded by a dynamic arithmetic coder.

3.4 Generation of Multiple Perspective Video

In our laboratory, we have developed a system based on disparity morphing technique to on-line generate multiple perspective stereo video by using the original two views and disparity maps. Since the image planes of original videos are approximately parallel. The position of each pixel in the generated video can be linearly interpolated with a viewpoint parameter s from the corresponding pixel positions and the corresponding disparities. The viewpoint parameter s defines the location of the viewpoint. When s is 0, the viewpoint is located at the original left eye viewpoint. and when s is 1, the viewpoint is located at the original right eye viewpoint. Figure 10 illustrates the meaning of parameter s . The generated multiple perspective views are shown in figure 11 when the viewpoint parameter s is from -1 to 2

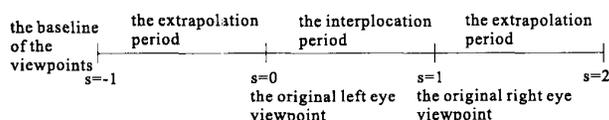


Figure 10: The meaning of viewpoint parameter s .

stepped by 0.2.

The generated video usually contains holes and occluded objects. In our system, the image holes can be eliminated by interpolating the values of adjacent pixels, and the object occlusion problem can be dealt with correctly by analyzing the disparity flows as described in the next paragraph.

Figure 12 shows an example of disparity flows. In this example, there is a background object from points p_1 to p_6 covered by a foreground object at points p_3 and p_4 . Thus the points p_3 and p_4 move faster than p_1 , p_2 and p_5 , p_6 . During the view interpolation procedure, for example $s = 0.5$, point p_3 and p_5 morph to the same image position. In traditional image morphing, the resulted pixel value is usually the weighted sum of the values of original points p_3 and p_5 , and this method generates ghost effects on the resulted image. In our system, the disparity of each corresponding point is examined to decide which point value should be output as the resulted pixel value. In this example, the disparity of point p_3 is larger than that of point p_5 , that means the object at point p_3 is closer to the camera than the object at point p_5 . Thus the pixel value of point p_3 , the point with larger disparity, is output as the resulted pixel value. By applying this algorithm, the views generation system can deal with the object occlusion problem correctly, and can generate clear and correct interpolated and extrapolated videos.

3.5 Postwarping

The postwarping procedure warps the generated parallel videos to desired viewing direction. The desired viewing direction is usually toward to the interested area of the scene. The postwarping procedure rotates the image planes and generates the desired views[8].

4 Experimental Results

This section shows the experimental results of applying the proposed algorithm on real stereo video. Figure 13 shows the 1st, the 30-th and the 55-th frames of stereo video "Dance". Figures 14, 15 and 16 are the generated interpolation and extrapolation images for the 1st, and 30-th, and 55-th frame pairs, respectively. The meaning of s in figures 14, 15 and 16 has been described in section 3.4 and figure 10.

For the first frame pair (figure 13(a), please notice that

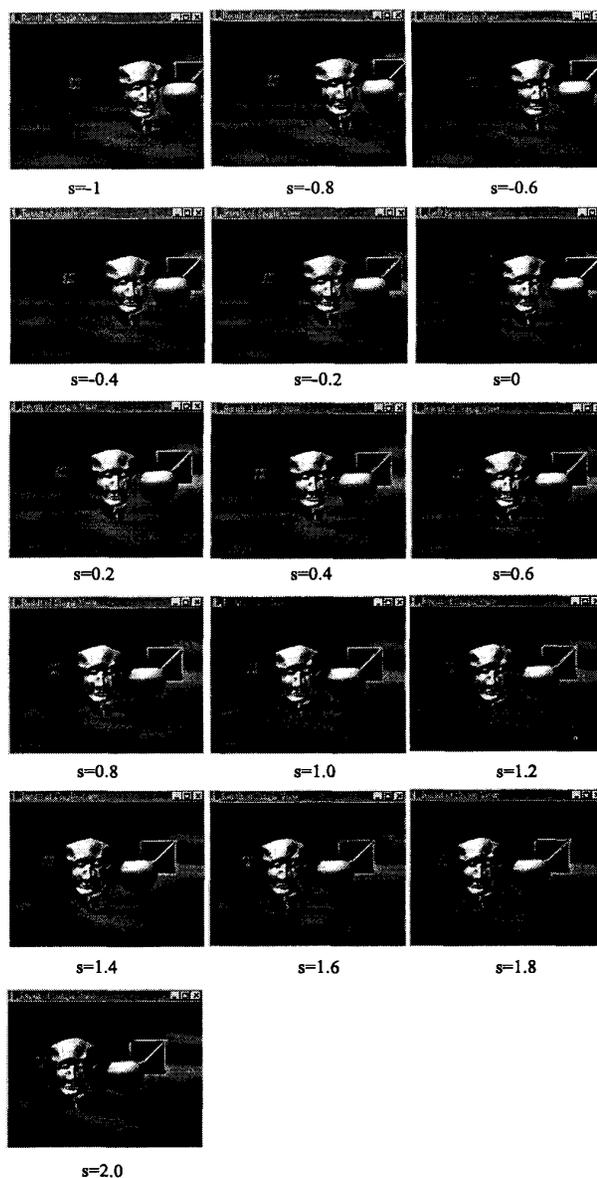


Figure 11: Generated views at different perspectives.

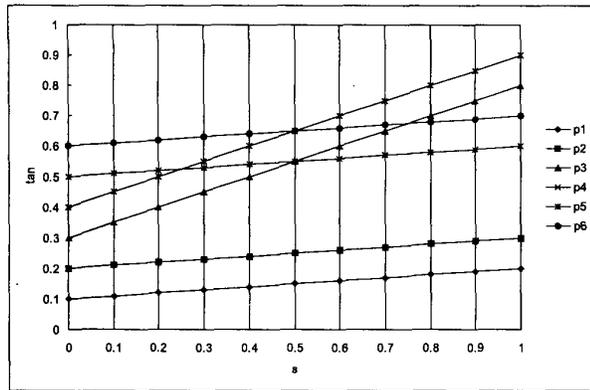


Figure 12: The example disparity flows of a correspondence scan line pair during the view interpolation procedure.

the head of the right dancer in right image is closer to the left boundary of the tent than that of the left image. As the generated views shown in figure 14, the distance between the head of the right dancer and the left boundary of the tent is larger when the viewpoint parameter s is smaller. For the 30-th frame pair (figure 13(b), please notice that the head of the right dancer in right image is farer from the right boundary of the tent than that of the left image. Examining the generated views shown in figure 15, the distance between the head of the right dancer and the right boundary of the tent is smaller when the viewpoint parameter s is smaller. For the 55-th frame pair (figure 13(c), please compare the relative positions between the waistband of the left dancer and the tent in both image. In all of the three examples, our systems can automatically generate correct views at different perspectives.

By combining with stereo glasses and head tracking system, users can see stereo video from arbitrary viewing directions. Figure 17 shows the user is using the multiple views generation system with an ultrasonic tracking system.

5 Conclusion and Future Work

We have developed a new method and system for generating multiple perspective stereo video from a two-view stereo video sequence. The system firstly generates disparity maps from original stereo video. The disparity maps is estimated based-on epipolar constraints, and smoothed by a refinement algorithm. The disparity maps generation procedure is highly automatic and highly accurate. The disparity maps as well as the original stereo video are compressed to reduce the storage size. Our system can on-line generate arbitrary perspective stereo video depending on the requirement of the user. Com-

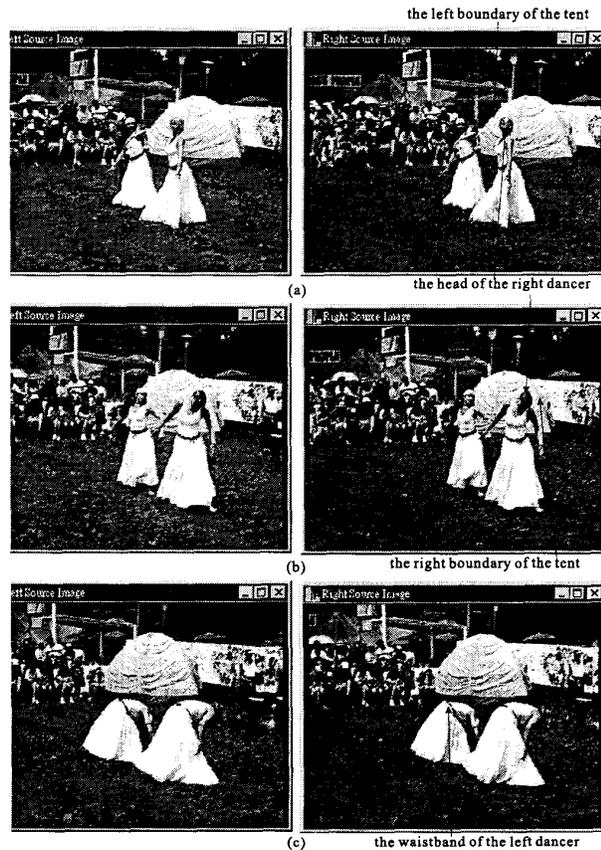


Figure 13: The original frames of stereo video "Dance", (a) the first frame pair, (b) the 30-th frame pair, and (c) the 55-th frame pair.

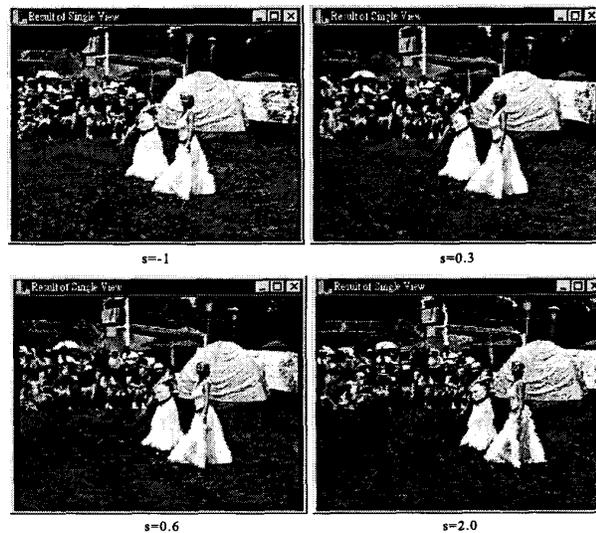


Figure 14: Generated views at different perspectives for the first frame pair of the stereo video "Dance".

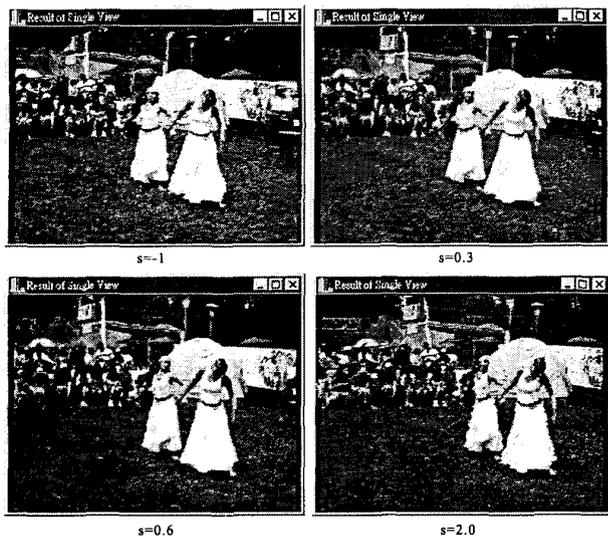


Figure 15: Generated views at different perspectives for the 30-th frame pair of the stereo video "Dance".

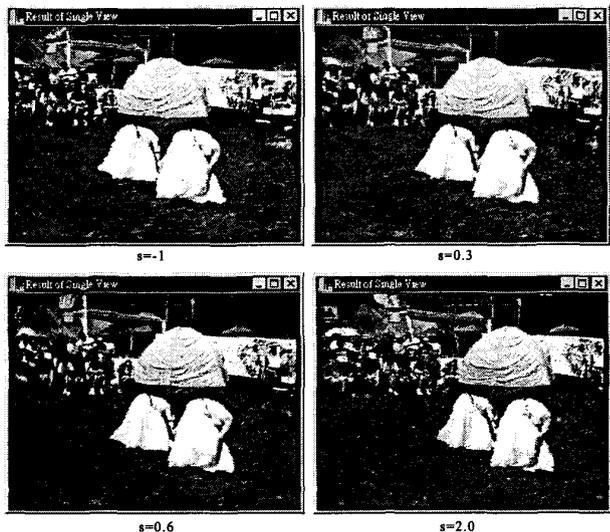


Figure 16: Generated views at different perspectives for the 55-th frame pair of the stereo video "Dance".



Figure 17: The user using our system to see the multiple perspective stereo video with a stereo eyeglasses and a head tracking system.

pared to the traditional method of capturing the multiple perspective video directly using multiple cameras, both the difficulty of capturing procedure and the amount of video data can be reduced significantly. Furthermore, since our system can interpolate or extrapolate the stereo video based on arbitrary viewing directions, the system can be easily applied to all kind of stereo monitor systems with different numbers of video projectors. The automatic correspondence estimation is a time-consuming procedure. How to increase both the estimation speed and accuracy is one of our future work.

Acknowledgments

This work is partially supported by Opto-Electronics & Systems Laboratory, Industrial Technology Research Institute, Hsinchu, Taiwan, ROC.

References

- [1] et al. S. Moezzi. Immersive video. In *Proceedings of the IEEE Virtual Reality Annual International Symposium*, March 1996.
- [2] et. Al. S. Moezzi. Reality modeling and visualization from multiple video sequences. *IEEE Computer Graphics and Applications*, pages 58–63, November 1996.
- [3] S. E. Chen. QuickTime VR — an image-based approach to virtual environment navigation. *SIG GRAPH '95*, pages 29–38, 1995.
- [4] H.-C. Huang Y.-P. Hung and S.-W. Shih. Panoramic stereo imaging with complete-focus views for virtual reality. In *Proceeding of Second Workshop on Real-Time and Media Systems, RAMS'96*, pages 405–410, Taipei, Taiwan, ROC, July 1996.
- [5] H.-C. Huang and Y.-P. Hung. Spisy: The stereo panoramic imaging system. In *RAMS'97, Third Workshop on Real-Time and Media Systems*, pages 71–78, Taipei, Taiwan, R.O.C., July 1997.

- [6] H.-C. Huang and Y.-P. Hung. Panoramic stereo imaging system with automatic disparity warping and seaming. *Graphics Models and Image Processing*, 60(3):196–208, May 1998.
- [7] T. Beiter and S. Neely. Feature-based image metamorphosis. *Proceedings SIGGRAPH '92*, pages 35–42, 1992.
- [8] S. M. Seitz and C. R. Dyer. View morphing. *Proceedings SIGGRAPH '96*, pages 21–30, 1996.
- [9] Q.-T. Luong and O. D. Faugeras. The fundamental matrix: Theory, algorithms, and stability analysis. *Inter. J. of Computer Vision*, 17(1):43–75, January 1996.
- [10] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. INRIA-report, No. 2927, July 1996.
- [11] R. I. Hartley. In defence of the 8-point algorithm. In *Fifth Inter. Conf. on Computer Vision*, pages 1064–1070, Massachusetts Institute of Technology, Cambridge, Massachusetts, June 1995.
- [12] H.-C. Huang, S.-H. Nain, Y.-P. Hung, and T. Cheng. Disparity-based—a new technique for image-based rendering. In *ACM Symposium on Virtual Reality Software and Technology 1998(VRST'98)*, pages 9–16, Taipei, Taiwan, ROC, November 1998.
- [13] R. Szeliski and H.-Y. Shum. Creating full view panoramic image mosaics and environment maps. *Proceedings SIGGRAPH '97*, pages 251–258, 1997.
- [14] M. Levoy and P. Hanrahan. Light field rendering. *Proceedings SIGGRAPH '96*, pages 31–42, 1996.
- [15] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. *Proceedings SIGGRAPH '95*, pages 39–46, 1995.
- [16] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. *Proceedings SIGGRAPH '96*, pages 43–54, 1996.
- [17] *MPEG standard draft ISO-IEC/JTC1 SC29 on 22*, November 1991.
- [18] *MPEG standard, ISO/IEC-11172: Coding of Moving Pictures and Associated Audio—for Digital Storage Media at up to About 1.5 Mbps*.
- [19] D. Le Gall. Mpeg: A video compression standard for multimedia applications. *Commun. ACM*, 34:47–58, April 1991.
- [20] B. L. Tseng and D. Anastassiou. Multi-viewpoint video coding with mpeg-2 compatibility. *IEEE Transactions on Circuits and Systems for Video Technology*, 6(4):414–419, 1996.
- [21] B. G. Haskell, A. Puri, and A. N. Netravali. *Digital Video: An Introduction to MPEG-2*. Chapman and Hall, 1997.
- [22] B. Furht et al. Design issues for interactive television systems. *IEEE Computer*, 28(5):25–39, 1995.
- [23] P. H. Kelly et al. An architecture for multiple perspective interactive video. In *ACM Multimedia 1995 Proceedings*, pages 201–212, CA, USA, November 1995.

Biographies



Ho-Chao Huang was born in E-Lan, Taiwan, ROC in 1968. He received the B.S. and Ph.D. degrees from the Computer Science and Information Engineering Department at National Taiwan University, Taipei, Taiwan, ROC in 1990 and 1995, respectively.

Since October 1995, he has been a Postdoctoral Research Fellow at the Institute of Information Science in Academia Sinica, Taipei, Taiwan, ROC. His research interests include image-based rendering, virtual reality, video signal processing, video object segmentation, and multimedia systems.

Ching-Che Kao received the B.S. and Ph.D. degree from the Computer Science and Information Engineering Department at National Taiwan University, Taipei, Taiwan, ROC in 1998. He is currently a graduate student of the Computer Science and Information Engineering Department at National Taiwan University. His research interests include Image-Based rendering, Computer Vision, and Stereo System.



Yi-Ping Hung received his B.S. in electrical engineering from National Taiwan University in 1982. He received an M.S. from the Division of Engineering, an M.S. from the Division of Applied Mathematics, and a Ph.D. from the Division of Engineering, all at Brown University, in 1987, 1988 and 1989, respectively. He then joined the Institute of Information Science, Academia Sinica, Taiwan, and became a research fellow in 1997. He served as the deputy director of the Institute of Information Science from 1996 to 1997, and received the Outstanding Young Investigator Award given by Academia Sinica in 1997. He has been teaching in the Department of Computer Science and Information Engineering at National Taiwan University since 1990, where he is now an adjunct professor.

Dr. Hung has published more than 70 technical papers in the fields of computer vision, pattern recognition, image processing, and robotics. In addition to the above topics, his current research interests also include visual surveillance, virtual reality, human-computer interface and visual communication.