http://dx.doi.org/10.6109/ijice.2011.9.6.780

3D Panorama Generation Using Depth-MapStitching

Seung-Il Cho, Jong-Chan Kim, Kyeong-Jin Ban, Kyoung-wook Park,
Kim Cheeyong, and Eung-Kon Kim, *Member, KIICE*

***Abstract*—As the popularization and development of 3D display makes common users easy to experience a solid 3D virtual reality, the demand for virtual reality contents are increasing. In this paper, we propose 3D panorama system using vanishing point locationbased depth map generation method. 3D panorama using depthmap stitching gives an effect that makes users feel staying at real place and looking around nearby circumstances. Also, 3D panorama gives free sight point for both nearby object and remote one and provides solid 3D video.**

***Index Terms*—3DPanorama, Depthmap, Stitching, Virtual Environment**

I. INTRODUCTION

3D virtual environment which makes users feel in real situation or location are applied in various area such as education, advertisement, home-shopping, army training or medical surgery. To increase user's immersion, this 3D virtual environment needs to provide almost the same quality of real situation. As recent development and deployment of 3D display makes people easy to experience 3D virtual reality, the demand for virtual reality content is exponentially increasing in many areas.

Existing 3D virtual environment is generally made using computer graphics modeling authoring tool. However, this method takes huge time and energy to build a realistic virtual environment. To create a virtual environment that is almost the same as real one, this method should estimates all objects in real environment, perform modeling process then locate the objects in virtual environment. Although, investing time and energy for creating the virtual reality, it hardly meets user's expectation of experience because the product from this method does not provide perfect virtual reality.

To solve this issue, many researches have proposed Image-based Virtual Reality System(IBVR) that builds 3D virtual environment using images of real environment. IBVR shows its excellency on VR as it provides realistic film such as picture of object or environment in a certain point based on information gathered from images of real environment. Additionally, as it does not need to existing 3D modeling process, its cost is low and generation speed of output video is faster. Therefore, it accommodate domestic researches on 3D solid video generation based on video synthesis and processing of images[1,2].

Complementing weak points of depth map generation method for the recovery of 3D space structure from 2D video, this paper proposes 3D panoramic system using loss point location-based depth map generation method. 3D panorama using depth map gives a panorama effect that a user feels as if he is in the point and looks around nearby circumstances. With our proposed system, users will get wide viewpoint of close ranged objects and remote background and realistic 3D video like high resolution images.

II. RELATED WORKS

A. Image mosaic

Image Mosaic is a synthesized video that takes scenery images by video capture equipment then combines the images for users to see all direction nearby himself.As it provides wide field of view(FOV) for one scene compared to single video captured by video capture equipment such as a camera, it is commonly used. Mosaic can be applied to various areas. For example, map making for submarine's navigation and for aerial photo or satellite photo and many other scientific purpose [3,4,5].

It is also a basic modeling process to create IBVR that is comprised of panoramic video mosaic by mixing series of videos. Recently, IBVR becomes a hot topic in computer vision area as well as computer graphics. Compared to virtual reality systems that are based on existing 3D model, IBVR provides not only better for supporting reality but also simpler rendering process. This method that is proposed for 2 D based video mosaic is not appropriate for building navigationenabled 3D virtual reality.

B. Panorama VR

Panorama VR is to show nearby sight at one's location. it gives an effect just like he feels in the place and it can reenact at certain days. To get this effect, panorama method is used. 360 angle image that is taken real world in one shot is called as Omni image or all direction image. Images that transform Omni image to cylinder, rectangular or globular shape are called as panoramic video.

Manuscript received September 9, 2011; revised October 13, 2011; accepted October 28, 2011.

Eung-Kon Kim is with the Department of Computer Science, Sunchon University, Suncheon, Jeonnam, 540-742, Korea (Email: kek@sunchon.ac.kr)

To generate panorama video, it divides 360 angle backgrounds into many components and takes each shot in which front image and rear image are overlapped about 1/3. Then it creates panorama effect by applying stitch algorithm to each image.

Representative technologies for this method are IPIX from Interactive Picture, Inc, HotMedia from IBM, QuickTime VR from Apple and RealityStudio from Live Picture and many other researches for panorama generation are currently held[6,7,8].

With panorama VR, it is easy to build virtual environment and provides video exactly the same as real world images, which enables application to present certain space such as apartment and advertisement of virtual gallery. As panorama VR only utilizes 2D image, it does not provide high solidity, limits viewpoint since users can see objects only from the place of taking and does not support various navigation functionality.

C. Perception Components of Depth

To generate 3D video, it is necessary to understand human perception of depth. There are many clues used to perception process that determines size, distance and depth of an object. These clues are categorized as single eye clue and both eyes clue. Single eye clue is when gathered information from single eye are used as clue and both eyes clue is when gathered information from both eyes are combined then used as one clue.

Single eye clues that affect depth perception are overlap, linear view, atmosphere view, aspect height, density change of quality, shadow and so on. Other single eye clues are such as motion parallax and accommodation. Overlap is to perceive one object, which screens another object, to be located in front. Linear view is to look both edges from two parallel lines to be narrower as they are convergent to one point in remote. Atmosphere view is to conceive contours of an object unclear and opaque in proportion to its remoteness and aspect height is for the object that is seen below user's view horizon to be seen farther than for the one that is seen above the horizon. Shadow makes an effect for an object to be seen more solid.

Motion parallax is one strong depth clue that it makes conception for the movement speed of an object to be faster in proportion to the observer's distance. For example, when travelling by train, trees that are nearby the observer looks move faster than trees in remote from the observer looks move slowly toward to the direction that the train goes to. Distance control is to control focal point of eye lens in accordance with the location of an object; video in retinas tenses the eye lens and movement of eye muscles change the focal point. When pointing close range, the eye lens become thicker and when pointing remote and wide range, it becomes thinner.

Both eyes clues are convergence and binocular disparity. Convergence is when an object become closer to the observer, eyeballs are convergent to inner direction to control eyes to trail the object and when an object is in remote, eyeballs are convergent to the center of the eye lens to narrow the range of eyesight. Fig 1 shows the principle of convergence. Both eyes clue means that when eyes of a man see real world, each eye gets different image. Binocular disparity comprises of positive parallax, negative parallax and zero parallax. Positive parallax is a case that observation point is farther than solid plane, whose parallax is bigger than 0 and it makes an effect of depth. Zero parallax is a case that observation point is in the same distance with that of solid plain and negative parallax, which occurs when line of sight crosses over and makes projection effect of solid plane, is a case that the observation point is closer than solid plane. Fig 2 shows these three types of binocular disparity. Both eyes clue is gained from the difference of image between right and left eye and is effective in close range. Single eye clue, however, can give more important solid information at more than 10m distance[9].

III. VR PANORAMA GENERATION ALGORITHM

Let $x=(x,y)$ which is a point from projectionof $X=(X,Y,Z)$, a point in 3 dimensional space that is taken by lens distortion free camera, to 2D image, the relationship between X and x is modeled as follow.

$$u=\begin{bmatrix}u\\v\\w\end{bmatrix}=K(RX+t)$$

$$=\begin{bmatrix}f & s & x_c\\0 & af & y_c\\0 & 0 & 1\end{bmatrix}\left(\begin{bmatrix}r_{11} & r_{12} & r_{13}\\r_{21} & r_{22} & r_{23}\\r_{31} & r_{32} & r_{33}\end{bmatrix}\begin{bmatrix}X\\Y\\Z\end{bmatrix}+\begin{bmatrix}t_1\\t_2\\t_3\end{bmatrix}\right), \quad (1)$$

$$x=\frac{u}{v},\quad y=\frac{v}{w} \quad (2)$$

u is a vector that represents x as Homogeneous Coordinates. K, as a Camera Calibration Matrix, shows the relationship between 3D point on camera coordinates and 2D point on image within which 3D point is projected. R and t, that represent Rotation Matrix and Translation Vector each, indicate the transformation relationship between world coordinates and camera coordinates. As we call parameters included in vector K as Internal Parameter of camera, f is Focal Length, Q is Aspect Ratio of pixel, s is Skew and (X_c, Y_c) is Principal Point.

$$u_1=KR_1X=KR_1R_{0.}^{-1}K^{-1}u_0=KR_{01}K^{-1}u_0=Hu_0 \quad (3)$$

Rotation Matrix and Translation Vector of camera before rotating are R_0 and t_0 respectively and those of camera after rotating are R_1 and t_1. When conforming projection center to zero point of world coordinates for convenience's sake, it becomes $t_0 = t_1 = 0$. If homogeneous vector of a point that is from the projection of certain point X at 3 dimensional space to I_0 (original image) and I_1 (rotated image) is each u_0 and u_1.

That is, as pixel correspondence relationship between two images that are from camera rotation, which is presented by a 3×3 matrix H, we call it Planar Homography. Considering that corresponding image point is the same even though multiplying any real number except 0 to Homogeneous Vector, H also can be defined regardless of its scale. Hence, Planar Homography is defined 8 parameters, not 9 parameters. Actually, as the number of parameter that is included in K from $H = HR_{01}K^{-1}$ is 5 and the number of parameter that is in R_{01} is 3, we can aware that independent parameters that are composed of H is 8. This Homography is called as 8Parameter Homography[10].

Given two images, finding H which is pixel correspondence relationship between the two images is Image Registration. In general, H can be calculated with more than 4 correspondent points but it takes caution to get more accurate H. Therefore, we assume that camera satisfies this normalized model.

TABLE I
CAMEARA CONDITION OF NORMALIZED MODEL

condition) - As pixels are generally square pixel, pixel's aspect ratio satisfies $a=1$. - As CCD cells that compose image are arranged on the square, Skew meets $s=0$. - As an assumption that optical axis of camera go through array center is safe, Principal Point (x_c, y_c) accords with image center.

Therefore, the number of independent parameter included in H on normalized model is 4(focal point distance f and three rotating angles within R_{01}, which we call as 4-Parameter Homography[11]. General rotation matrix R is divided as follow.

$$R = R(\psi)R(\theta)R(\varphi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos\varphi & 0 & \sin\varphi \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

In formula(4), $R(\psi), R(\theta), R(\varphi)$ are respective matrix that present rotation transformation from each X, Y, Z axis and ψ, θ, φ are called Euler Angle. Given Planar Homography H in normalized model, calculating K and R are as shown below. Matrix K can be divided as follow.

$$K = K_c K_f = \begin{bmatrix} 1 & 0 & x_c \\ 0 & 1 & y_c \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

Substituting to formula(3)

$$K_f = RK_f^{-1} = K_c^{-1}HK_c = \hat{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (6)$$

Therefore

$$R = \begin{bmatrix} h_{11} & h_{12} & h_{13}/f \\ h_{21} & h_{22} & h_{23}/f \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (7)$$

As R is Orthogonal Matrix, it satisfies following formula.

$$h_{11}^2 + h_{12}^2 + \frac{h_{13}^2}{f^2} = h_{31}^2 f^2 + h_{32}^2 f^2 + h_{33}^2 \quad (8)$$

After calculating f from formula(8), it gets R by substituting to formula(7).

Camera models described above do not consider distortion of its lens. As it utilizes wide angle lens to get 360 angle panorama video, it is necessary to consider lens distortion to apply algorithm to wider area. (x_d, y_d) is coordinates of pixel in original image and (x_u, y_u) is corresponding coordinates of pixel in image that removes lens distortion. As (x_c, y_c) is the center coordinates, given (x_d, y_d), (x_u, y_u) is calculated as follow[12].

$$\begin{aligned} x_u &= (x_d - x_c)(1 + k_1 r_d^2 + k_2 r_d^4 + \ldots) + x_c, \\ y_u &= (y_d - y_c)(1 + k_1 r_d^2 + k_2 r_d^4 + \ldots) + y_c, \\ r_d &= \sqrt{(x_d - x_c^2) + (y_d - y_c^2)} \end{aligned} \quad (9)$$

In this moment, k_1, k_2 are radial distortion coefficients.

$$\begin{aligned} x_u &= (x_d - x_c)(1 + k r_d^2) + x_c, \\ y_u &= (y_d - y_c)(1 + k r_d^2) + y_c, \end{aligned} \quad (10)$$

To the contrary when (x_u, y_u) is given, (x_d, y_d) is as follow.

$$x_d = \frac{x_u - x_c}{1 + k r_d^2} + x_c, \; y_d = \frac{y_u - y_c}{1 + k r_d^2} + y_c \quad (11)$$

In formula(12),

$$r_d^2 = h(k) + \frac{1}{9h(k)k^2} - \frac{2}{3k}, \quad (12)$$

$$h(k) = \left(\sqrt{\left(\frac{1}{27k^3} + \frac{r_u^2}{2k^2}\right)^2 - \frac{1}{729k^6}} + \left(\frac{1}{27k^3} + \frac{r_u^2}{2k^2}\right)\right)^{\frac{1}{3}}, \quad (13)$$

$$r_u^2 = (x_u - x_c)^2 + (y_u - y_c)^2 \quad (14)$$

IV. 3D PANORAMA SYSTEM USING DEPTH MAP

If generated video from computer vision and computer graphics technologies and virtual environment related technologies that are to interact users are effectively converged, VR panoramic 3D virtual environment can be manufactured. Fig 1 shows the concept of system that we propose in this paper.

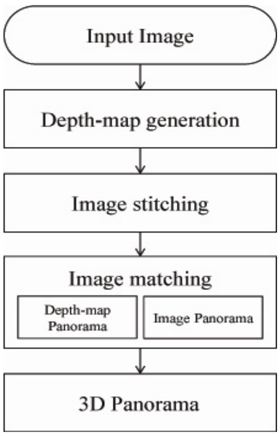


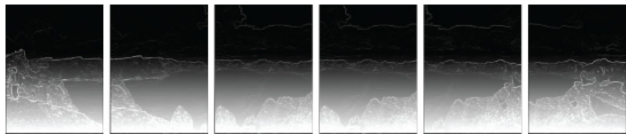
Fig. 1. System architecture of VR panoramic system using depth map.

Generation of VR panorama image using depth map is to build wide angle panorama image that include 360 angle view by weaving multiple videos that are taken by rotating camera in a fixed point. This system gives an effect that user feels as if he looks all direction in the shot point. This can be done using just one tripod, needless to have any special equipment. Even images that are taken by camera which is not horizontally placed when taking pictures or videos can be stitched. By automatically revising brightness difference between input images, it creates sphere projection panorama image.

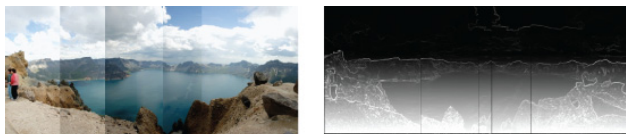
In image processing area, image stitching method is a task that manipulates taken videos in series angles as outstretched format. Depth map extraction applies median filter to remove noise in video. It extracts edges from pre-processed input video by applying canny edge detection method then traces vanishing lines by Hough transform. It designates vanishing point through cross points of extracted vanishing lines.

Generation of depth map is done with three steps. Edge extraction, the first step, is a pre-process for input video, applying median filter to reduce noise and canny edge to detect geometric characters. Vanishing point and vanishing lines generation step, the second step, extract straight lines and their crossing point by hough transform and estimate the location of vanishing point by the range of defined vanishing point. Vanishing point estimation and depth map generation, the last step, estimate the location of vanishing point by the range of defined vanishing point location, setup depth level and generate depth map based on the location of vanishing point.Fig 2.shows the process of creating solid VR panoramic video by stitching each video based on extracted depth map, creating depth map panorama and matching images.

(a) Input images

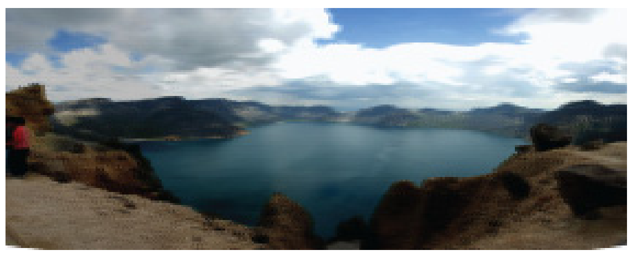


(b) Depthmap extraction



(c) Image/Depthmap stitching

(d) Image/ Depthmap panorama



(e) 3D panorama

Fig. 2. Realization of 3D panoramic video based on depthmap.

V. CONCLUSIONS

To complement weak point of depth map generation method from 2 dimensional video to 3 dimensional space structure recovery, this paper proposes VR panoramic system using depth map generation method which is based on the location of vanishing point in video in which existing researches has been unable to define.

Based on computer vision technology and computer graphics technology, depth map based 3D panoramic virtual environment system can create realistic 3D panoramic virtual environment if manufactured video and user interaction can be effectively convergent.

Using our proposed system, it can overcome the limit of existing 2D panoramic technology and give a more natural navigation so that user can experience more reality and better qualify of immersion. To generate this type of solid video, not only nearby point but also remote point needs to find 3D depth information.

Our future issue is to develop VR authoring tool so that low coat VR content generation can be realized.

REFERENCES

[1] D. Burschka et al,."Recent Methods for Image-Based Modeling and Rendering,"IEEE Virtual Reality, pp. 299, Mar, 2003.

[2] H. K. Shum, S. B. Kang, and S. C. Chan, "Survey of image-based representations and compression techniques," IEEE Trans. On Circuits and Systems for Video Technology, Vol. 13, No. 11, pp.1020-1037, Nov, 2003.

[3] R. Szeliski, "Image mosaicing for tele-reality applications," Proc. of the IEEE Workshop on Applications of Computer Vision, pp. 44–53, 1994.

[4] X. Xu, and S. Negahdaripour, "Vision-based motion sensing from underwater navigation and mosaicing of ocean floor images," Proc. of the MTS/IEEE OCEANSConference, vol. 2, pp.1412–1417, 1997.

[5] R. Garcia, X. Cufí, X. Muñoz, L. Pacheco, J. Batlle, "An image mosaicking method based on the integration of grey level and textural features," IX SimposiumNacional de Reconocimiento de Formas y AnálisisdeImágenes (SNRFAI), Benicassim, Castellón, 2001.

[6] S. E. Chen, "Quicktime VR – an image-based approach to virtual environment navigation," in ACM Computer Graphics (SIGGRAPH'95), pp. 29-38, Aug. 1995.

[7] L. Juan and O. Gwun, "SURF applied in panorama image stitching," 2010 2nd International Conference on Image Processing Theory Tools and Applications (IPTA), pp. 495-499, July, 2010

[8] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," International Journal of Computer Vision, Vol. 74, No. 1, pp. 59-73, 2007.

[9] K. J. Ban, J. C. Kim and E. K. Kim," An Object Representation System Using Virtual Space Coordinates," KIMICS, Vol.8, No.4, Aug, 2010.

[10] R. Szeliski, "Video Mosaics for Virtual Environments," IEEE Computer Graphics and Applications, Vol. 16, No. 2, pp. 22-30, 1996.

[11] R. Szeliski and H. Y. Shum, "Creating Full View Panoramic Image Mosaics and Environment Maps, "Proc. SIGGRAPH, 1997.

[12] N. Chiba, M. Minoh, and H. Kano, "Feature-Based Lens Distortion Correction for Image Mosaicing, "Proc. IAPR Workshop on Machine Vision Applications (MVA), 2000.

Seung-Il Cho He reveived his B.S degree from Chosun University in 1992,His M.S degree from department of computer education, Sunchon National University in 2001,His Ph.D degree from department of computer science, Sunchon National University in 2008.His current research interests are computer graphics, multimedia and image processing.

Jong-Chan Kim He received his B.S. degree from Sunchon National University in 2000, his M.S degree from department of computer science, Sunchon National University in 2002, his Ph.D. degree from department of computer science, Sunchon National University in 2007.His current research interests are image processing, computer graphics, AR, games and HCI.

Kyeong-Jin Ban He received his B.S. degree from Sunchon National University in 2003, his M.S degree from department of computer science, Sunchon National University in 2005, his Ph.D degree from department of computer science, Sunchon National University in 2010.His current research interests are computer graphics, image processing, HCI, VR and AR.

Kyoung-Wook Park He received his B.S. degree from Sunchon National University in 1997, his M.S degree from department of computer science and statistics, Chonnam National University in 1999, his Ph.D. degree from department of computer science, Chonnam National University in 2004.His current research interests are parallel and distribution processing, graph theory, theory of computation, bioinformatics and Augmented Reality.

KimCheeyong He received the Ph.D. degree in Computational Physics from Inje University,Republic of Korea, in 2000.He worked at Inje University, Busan College from 1991~2002and at Dongseo University from 2003-2005.In 2011, he joined the Dong-Eui University, where he is an associate professor.His research interests are in 3D Animation, Multimedia Design and VR,Contents Design, Chaos & Fractal Design.

Eung-Kon Kim He received B.E. degree Cho-sun University in 1980, his M.E. degree from depar-tment of computer engineering, Hanyang University in 1986, his Ph.D degree form depart-ment of computer engineering, Chosun University in 1994, In 2011, he joined department of computer science, the Sunchon National University, where he is a professor. His current research interests are image processing, AR,computer graphics, multimedia and HCI.