International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.9 (2016), pp.117-128
http://dx.doi.org/10.14257/ijmue.2016.11.9.13

A Study on 3D Panorama System Using Depth-map Generation Techniques

Jong Chan Kim, Oh Hoon Cho and Eung Kon Kim

Department of Computer Science Sunchon National University, 540-712, 315 Maegok, Suncheon Jeonnam, Republic of Korea
{seaghost@sunchon.ac.kr, worldnet@nate.com, kek@sunchon.ac.kr}

Abstract

The point is a vanishing point and the lines which meet at the point are vanishing lines. Depth-map can be generated to restore a 3D space structure from the 2D images using the vanishing point. The research described in this paper compensates for the disadvantages of depth-map generation techniques that need to restore a 3D spatial structure from a 2D image. In addition, a 3D panorama system is proposed. The 3D panorama system using depth-map generation techniques creates a 3D effect in which users feel that they are standing at the shooting location and looking around. The 3D panorama system provides a free time for nearby objects and distant background in a wide area, and it creates realistic 3D images such as constant high-quality pictures. The 3D panorama system will offer the sense for the real and the immersion as it gets over the limit of 2D panorama technology and supports users with a natural navigation. The 3D panorama image using depth-map is worth noticing since it shows the possibility to use the construction of image based virtual system.

Keywords*: Computer graphics, Vanishing point, 3D panorama, Depth-map, Stitching*

1. Introduction

The people who research computer graphics (CG) have worked to improve given 3D models to be more realistic. By those efforts, 3D graphics technique reached considerable level. Even it is possible to produce a film by using only computer graphics techniques. The motion applied 3D graphics is based on real world. In computer vision area, the research to make more realistic 3D model or virtual environment has actively progressed. The conventional virtual environments are formed by CG modeling method. However, using a modeling software to make realistic virtual environment needs lots of times and efforts [1,10-11].

Mosaic image, panorama VR (virtual reality), and TIP(tour into the picture) methods have been studied to produce image-based VR systems for building a virtual environment. The mosaic image method is used to compose multiple images that are taken from the actual viewpoint of the user. Although this method has the advantage of easy production, it is not appropriate for organizing a 3D virtual environment.

Panorama VR is used for easily organizing a virtual environment and for expressing a particular space such as a model home or virtual exhibition owing to its advantage of providing real-world images. However, panorama VR lacks three dimensions owing to its use of 2D images, and the time depicted is limited as only objects at the time of the shooting location are viewed; in addition, it does not provide various navigation functions. TIP adds to the advantage of panorama VR by applying navigation functions, and its three-dimensionality allows the creation of new images by adding a 3D model. At present, both at home and abroad, studies are being carried out for creating virtual environments using TIP techniques that make users feel free time, and that include navigation functions and three-dimensionality [2-5]. TIP to date techniques are targeted to specific objects or

ISSN: 1975-0080 IJMUE

confined to a 2D image, so depth are less for close objects and long distance in the background.

In this research paper, we proposed a new 3D panorama system by using depth creation using vanishing point. It can make up the weak point of former depth creation system to restore 2D images into 3D images. 3D panorama system using depth creation can give a cubic effect to its user as they are on the shooting point. The system can provide wide range of free viewpoint, realistic and cubic effect from a short distance object to long distance scenery. To compose realizing image including image processing and interpreting, we can produce multiple scale 3D virtual environments for user to user interaction.

2. Related Works

2.1. Depth Creation Using Vanishing Point

Vanishing points are powerful clues for depth perception in indoor or outdoor environment with artificial structures. A vanishing point corresponds to the point at the farthest distance from observers. Human beings can estimate a vanishing point using the geometrical components of a single-eye image, and perceive relative depth based on the location of the vanishing point and the viewpoint of the observer.

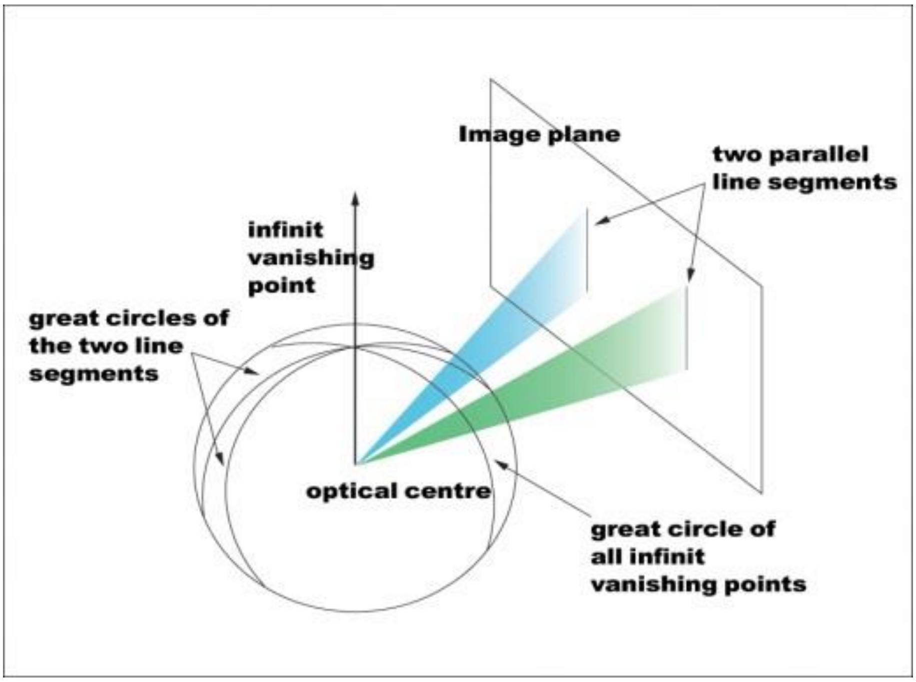

Figure 1. Gaussian Sphere for Detecting Vanishing Points

Vanishing point is one of the depth clues used to solve the problem of restoring the 3D spatial structure from 2D images. Images that contain artificial structures such as buildings, roads, and railroads consist of geometrical elements such as straight lines, and include many parallel or orthogonal straight lines. The distance between two parallel lines in a 3D space becomes shorter as they get farther from the observer in a projected image on a 2D plane and ultimately meet at one point. This point is called a vanishing point and the lines that meet at the vanishing point are called vanishing lines. Vanishing points are very powerful clue for depth in outdoor images that contain architectural structures such as buildings and roads and indoor images such as corridor and room. As vanishing points are generally on the horizon, they represent the farthest point from the bottom of the image. Thus, it is possible to estimate the relative depth using vanishing points. As a method for detecting vanishing point using straight lines, the Gaussian sphere is used [6-7]. Figure 1 illustrates the detection of vanishing point using the Gaussian sphere.

2.2. Panorama VR and Image Mosaic

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. A 360 angle image that is taken real world in one shot is called as Omni image or all direction images. Images that transform Omni images to a cylinder, a rectangular or a globular shape are called panoramic videos. To generate a panoramic 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 Reality Studio from Live Picture and many other researches for panorama generation are currently held. 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 a certain space such as an apartment and an advertisement of virtual gallery. As panorama VR only utilizes a 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 functionalities.

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 the user. 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 purposes. 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 2D based video mosaic is not appropriate for building navigation enabled 3D virtual reality [8-9].

3. 3D Panorama Generation Algorithm

Let $x = (x, y)$, which is a point from the projection of $X = (X, Y, Z)$, be a point in 3D space that is taken by a lens-distortion-free camera. For a 2D image, the relationship between X and x is modeled as follows.

u is a vector that represents x as a homogeneous coordinate. K is a camera calibration matrix that shows the relationship between a 3D point of the camera coordinates and a 2D point on the image upon which the 3D point is projected. R and t represent a rotation matrix and translation vector, respectively, and indicate the transformation relationship between world coordinates and the camera coordinates. As we call the parameters included in vector K the internal parameters of the camera, f is the focal length, Q is the aspect ratio of the pixels, s is the skew, and (X_c, Y_c) is the principal point.

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

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

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

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

The rotation matrix and translation vector of the camera before rotating are R_0 and t_0, respectively, while those of the camera after rotating are R_1 and t_1. For convenience, when conforming the center of a projection to the zero point of the world coordinates, we obtain $t_0 = t_1 = 0$. 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, we refer to the pixel relationship between two images from a camera rotation as planar homography; this relationship is represented by a 3×3 matrix H. Considering that the corresponding image point is the same when multiplying any real number except 0 with the homogeneous vector, H can also be defined regardless of its scale. Hence, planar homography is defined using eight parameters, not nine. In fact, as the number of parameters included in K from $H = HR_{01}K^{-1}$ is five and the number of parameters in R_{01} is three, we can see that the number of independent parameters of H is eight. This type of homography is called 8-parameter homography.

Table 1. Camera Condition in Normalized Mode

(condition)

- As pixels are generally squares, the pixel aspect ratio satisfies $a = 1$.
- As the CCD cells that compose an image are arranged on the square, the skew is $s = 0$.
- As the assumption that the optical axis of the camera goes through the array center is a safe one, the principal point (x_c, y_c) is in accordance with the image center.

Finding H, which indicates the pixel relationship between two images, is called image registration. In general, H can be calculated using more than four correspondent points, but greater caution should be used to obtain a more accurate value of H. Therefore, we assume that the camera satisfies this normalized model.

Therefore, in the normalized model, the number of independent parameters included in H is four (focal point distance f and three rotating angles within R_{01}; we refer to this as 4-parameter homography). General rotation matrix R is divided as follows.

$$R = R(\psi)R(\theta)R(\varphi)$$

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

In equation (4), $R(\psi), R(\theta), R(\varphi)$ are the respective matrices that present the rotation transformation from each X, Y, Z axis, and ψ, θ, φ are the Euler angles. Given planar

Homography H in a normalized model, the calculations of K and R are conducted as shown below. Matrix K can be divided as follows.

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

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

Substituting with equation (3),

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

Therefore,
As R is an orthogonal matrix, it satisfies the following equation.

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

After calculating f from equation (8), R is obtained by substituting f in equation (7). The camera models described above do not consider lens distortion. As a wide-angle lens is used to obtain a 360° panorama video, it is necessary to consider lens distortion when applying the algorithm to a wider area. (x_d, y_d) are the coordinates of a pixel in an original image, and (x_u, y_u) are the corresponding coordinates of a pixel in an image in which lens distortion is removed. As (x_c, y_c) are the center coordinates, given (x_d, y_d) , (x_u, y_u) are calculated as follows:

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

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

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

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

Contrary to when (x_u, y_u) are given, (x_d, y_d) are as follows:
As equation (12),

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

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

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

4. 3D Panorama System Using a Depth-Map

Author names and affiliations are to be centered beneath the title and printed in Times New Roman 12-point and non-boldface type. Multiple authors may be shown in a two or three-column format, with their affiliations below their respective names. Affiliations are centered below each author name, italicized, not bold. Include e-mail addresses if possible. Follow the author information by two blank lines before main text. 3D panorama image generation using a depth- map can create a high-quality depth-map panorama image with only a tripod, without the need for special equipment such as a rotator. It is also possible for the camera to attach some pictures shot non-horizontally. The camera automatically compensates for the difference in brightness between the input pictures, and generates a cylindrical or spherical panorama image projection. Figure 2 shows the shooting direction of the camera rotation.

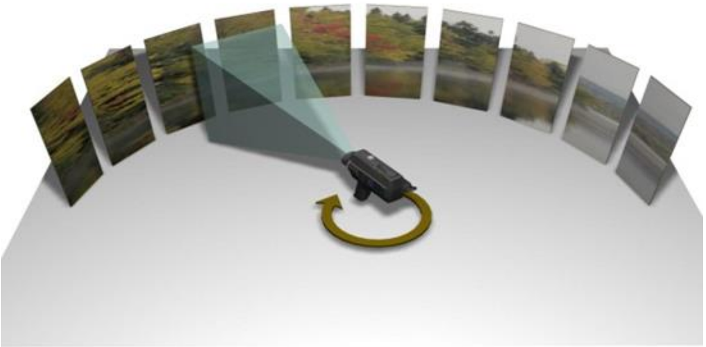

Figure 2. Shooting Direction of the Camera Rotation

The panorama image of a wide angle including 360 degrees attaching multiple image shot by rotating the body is generated. The generated depth-map panorama system gives the users the feeling that they are standing at the shooting location and looking around. Figure 3 shows the configuration of the panorama system with the proposed depth- map.

In the detection of a depth- map, a median filter is applied to remove noise from the image. It detects an edge on the input image through canny edge detection using pre-processing, and the vanishing points through a Hough transformation. The vanishing points are designated through the intersection points of the detected vanishing lines. As vanishing points appear in a 3D space and straight lines are more distant on a 2D fluoroscopy image, the gap is narrower, and the end points are thus the vanishing points. Straight lines meeting at the vanishing points are the vanishing lines.

The location scope of the vanishing points responding between images and vanishing points is defined because vanishing points are required to be classified into the depth phases depending on the image location. The standard-depth step on a plan is created based on the location scope of the defined vanishing points, and a depth- map is generated by linearly interpolating adjacent vanishing points.

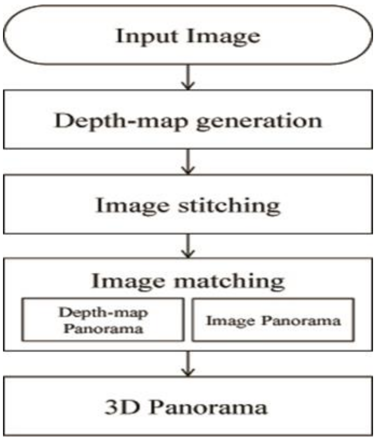

Figure 3. Configuration of the Panorama System With a Depth-map

Depth-map generation is divided into three steps. The first step, i.e., edge detection, reduces noise by applying a median filter as a pre-processing step on the input image, and detects the geometric characteristics by applying a canny edge detection algorithm. In the second step, i.e., a vanishing line and vanishing point generation, straight lines are detected through the application of a Hough transformation. In addition, the intersection points of the straight lines are obtained, and the location of the vanishing points are estimated through the location scope of the vanishing points. In the last step, i.e., the location estimation of the vanishing points and depth-map generation, the location of the vanishing points is estimated through the location scope of the vanishing points, set depth step, and a depth-map is ultimately generated on the basis of the location of the vanishing points.

This experiment is operated by the program materialized in visual C++ 2010 using OpenCV 2.3 library in Intel i5 1.83Ghz, 4GB, Windows 7 environment. Input data are used from image downloaded from 1MB pixel web camera and taken from 800MB pixel photograph and lots of JPEG image files.

In this paper, 3D panorama generation techniques using the proposed depth- map were studied for an outdoor image shot in both left and right directions for 180° on the basis of the cameras used. Figure 4 and 5 shows a comparison between the 3D panorama generation techniques using the proposed depth-map and the conventional depth-map generation techniques; 5 images shot at intervals of 40° degrees with process generation were used in the comparison; the accuracy of the depth-map was also compared.

Table2 shows the result of each method of required time. Proposed method needs about twice times more than the conventional method. Because it progresses one stage more. Table 3 shows that the depth-map for the middle area of the image has an accuracy of more than 85% in both the conventional and proposed methods. Conventional depth-map generation techniques do not generate an accurate depth-map for the left and right areas of the image. 3D panorama generation techniques using the proposed depth-map show a relatively high level of accuracy.

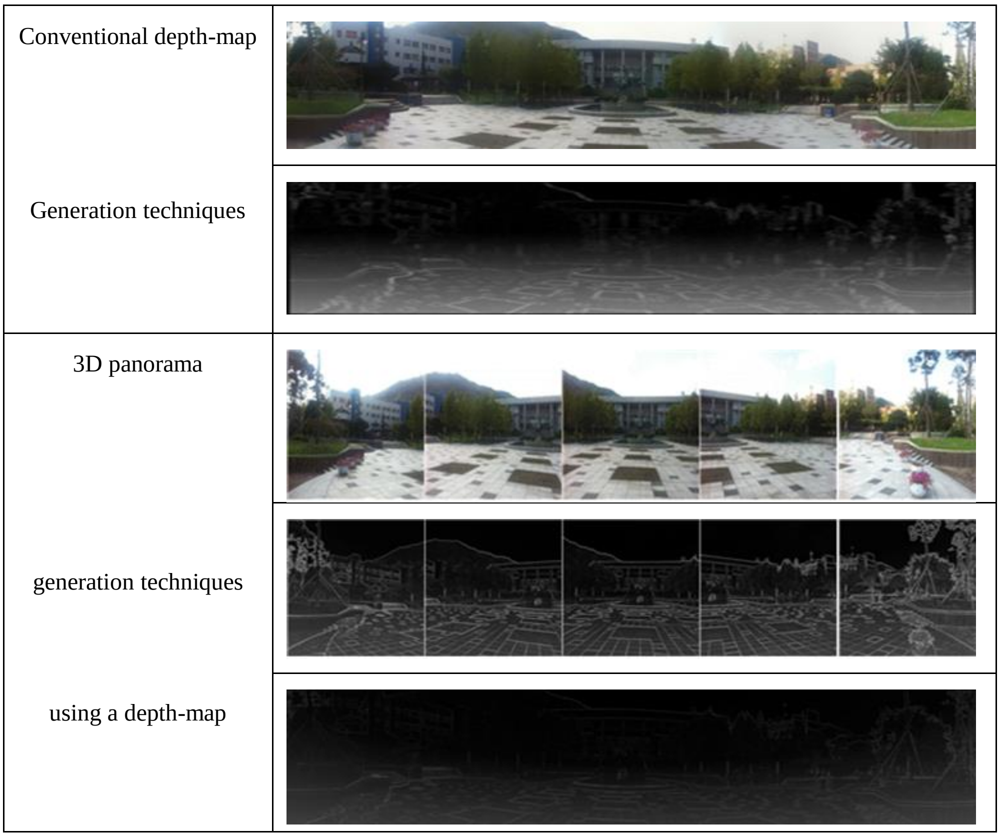

Figure 4. Panorama Depth-Map Generation Comparisons

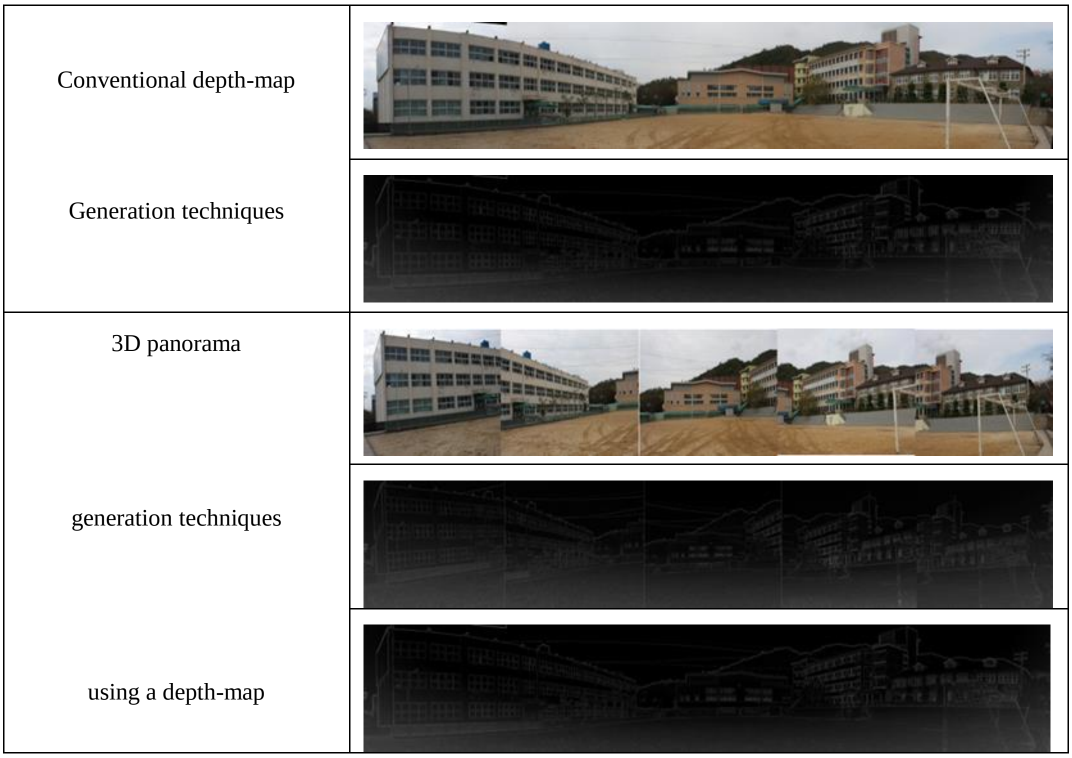

Figure 5. Panorama Depth-Map Generation Comparisons II

Table 2. Required Times

Depth-map generation techniques	Process stage	Time (msec)	Total time
Conventional depth-map generation techniques	Edge detection	86	320
	Line detection	68	
	Estimation vanishing point	9	
	Depth-map stage	76	
	Depth-map generation	81	
3D panorama generation techniques using a depth map	Edge detection	90(18x5)	660
	Line detection	245(49x5)	
	Estimation vanishing point	20(4x5)	
	Depth-map stage	190(38x5)	
	Depth-map generation	85(17x5)	
	Panorama depth-map coordination	30(30x1)	

Table 3. Comparison of Depth-Map Accuracy on a Panorama Image

Depth-map generation techniques	Match rate(%)			
		left	middle	right
Depth-map generation techniques using an estimation of the location of the vanishing points	Image Ⅰ	37	85	40
	Image Ⅱ	29	88	30
3D panorama generation techniques using a depth-map	Image Ⅰ	76	88	80
	Image Ⅱ	82	85	79

5. Conclusion

Computer vision system contributes a lot to user interface, real time object recognition, realizing media system to be developed. The various researches for 3D contents including cubic effect, immersion and sense of reality are actively progressed in domestic and foreign.

3D panorama generation techniques using depth-map has a cubic effect of looking around standing on shooting location to users. In 3D panorama system, it provides free viewpoint of short-distance object and long-distance scenery and realistic 3dimension stereoscopy like high quality picture in 3D virtual environment. To remedy the shortcomings of generation techniques of depth-map in 2D image, we proposed a 3D panorama system using generation techniques of depth-map by the point of vanishing point in an image. 3D panorama generation techniques using a depth-map needs more stage, so it takes 2.5 times more than conventional techniques. But its accuracy is improved about 2.5 times in left and right side of the image.

By using this paper, we can overcome the limit of conventional 2D panorama techniques and provide more realism and immersion by giving natural navigation to its users. It becomes the subject of interesting by proposing the possibility of constructing image based virtual reality system. 3D panorama generation techniques using a depth-map can produce realizing 3D panorama by combining effectively computer vision technique including image processing, interpreting technique and virtual environment method for interaction user-to-user.

Acknowledgements

This research was financially supported by the Ministry of Education (MOE) and National Research Foundation of Korea (NRF) through the Human Resource Training Project for Regional Innovation (No. 2013H1B8A2032217).

References

[1] 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, **(2003)**, pp. 1020-1037.

[2] R. Szeliski, "Image Mosaicing for Tele-reality Applications", Proceeding of the IEEE Workshop on Applications of Computer Vision, **(1994)**, pp. 44-53.

[3] X. Xu, and S. Negahdaripour, "Vision-based Motion Sensing from Underwater Navigation and Mosaicing of Ocean Floor Images", Proceeding of the MTS/IEEE OCEANS Conference, vol. 2, **(1997)**, pp. 1412–1417.

[4] R. Garcia, X. Cufí, X. Muñoz, L. Pacheco and J. Batlle, "An Image Mosaicking Method Based on the Integration of Grey Level and Textural Features", IX Simposium Nacional de Reconocimiento de Formas y Análisis deImágenes (SNRFAI), Benicassim, Castellón, **(2001)**.

[5] S. E. Chen, "Quicktime VR – an Image-based Approach to Virtual Environment Navigation", in ACM Computer Graphics (SIGGRAPH'95), **(1995)**, pp. 29-38.

[6] J. A. Shufelt, "Performance Evaluation and Analysis of Vanishing Point Detection Techniques", Proceeding ARPA Image Understanding Workshop, **(1996)**, pp. 1113-1132.

[7] K. J. Ban, J. C. Kim and E. K. Kim, "An Object Representation System Using Virtual Space Coordinates", KIMICS, vol. 8, no. 4, **(2010)**, pp. 411-415.

[8] M. Brown and D. G. Lowe, "Automatic Panoramic Image Stitching using Invariant Features", International Journal of Computer Vision, vol. 74, no. 1, **(2007)**, pp. 59-73.

[9] N. Chiba, M. Minoh, and H. Kano, "Feature-Based Lens Distortion Correction for Image Mosaicing", Proceeding IAPR Workshop on Machine Vision Applications (MVA), **(2000)**, pp. 28-30.

[10] Y. Muto, K. Okano, and S.Kusumoto, "A Visualization Technique for Unit Testing and Static Checking with Caller-Callee Relationships", JoC, **(2011)**, vol. 2, no. 2, pp. 1-8.

[11] A. M. Elmisery, and D. Botvich, "Enhanced Middleware for collaborative Privacy in IPTV Recommender Services", JoC, vol. 2, no. 2, **(2011)**, pp. 33-42.

Authors

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 the Department of Computer Science, Sunchon National University in 2007. He's a senior research professor of the Automation and System Research Institute at Seoul National University in 2013. His current research interests are image processing, computer graphics, digital clothing, AR, games and HCI.

Oh Hoon Cho, He received the B.S. degree from Korea National Open University, Korea, in 2006, his M.S. degree from Department of Computer Science, Sunchon National University, Korea, in 2013. He is currently a Ph.D. student in Computer Science at the Sunchon National University in Korea. His current research interests include augmented reality, image processing and computer graphics.

Eung Kon Kim, He received the B.S. degree from Chosun University, Gwangju,, Korea, in 1980. His M.S degree was from Department of Electronics in Hanyang University in Seoul, Korea in 1987. His Ph.D. degree was from Chosun University in Gwangju, Korea in 1992. His current research interests are computer vision, virtual/augmented reality, image processing, and computer graphics. Currently he is a professor in department of computer engineering, Sunchon National University in Korea.