Acta Technica 62 No. 1B/2017, 457–468

Virtual reality remote control system based on image[1]

Lin Gan[2], Jun Du[2], Xingcong Fu[3]

Abstract. In order to improve the effectiveness of remote control, we designed a virtual reality remote control system based on image. We implanted virtual reality technology to improve the remote control performance of unmanned vehicles. The system supports viewpoint transformation and overcomes the limitation of observation field restriction in traditional vehicle remote control. The simulation results show that the system provides a theoretical and technical basis for the application and development of 3D visualization and remote display system of unmanned vehicle. Also, the new tracking system based on image can support scene roaming at any point of view, which overcomes the restriction of the limited observation field of vision in the traditional vehicle remote control mode, and meets the demand of unmanned vehicle remote display. The system we designed solved problems of the camera limited field view and cannot obtain 3D environment information.

Key words. Unmanned vehicle, virtual reality, 3D terrain, texture mapping, image stitching
.

1. Introduction

With the rapid development of artificial intelligence, computer and sensor technology, a mobile robot, which uses sensors to sense its own state and external environment, has gradually entered a substantive application phase. The unmanned vehicle is one of the special intelligent mobile robots. It can complete the continuous and independent movement under a variety of complex roads or environmental conditions. It has been widely used in people's daily life, as well as in the national military operations and many other fields [1]. Essentially, the intelligent system of unmanned vehicle is a complex whole with many functions, such as integrated route planning, environment perception and moving direction control. At present, the related research has not carried on the thorough analysis to the remote-control op-

[1]The authors acknowledge the National Natural Science Foundation of China (Grants: 51578109 and 51121005).

[2]North China University of Science and Technology, Tangshan, Hebei 063210, China

[3]Beijing Huilong Technology Co., Ltd., BeiJing, 101200, China

erating system based on the scene and the image. In order to realize remote control of unmanned vehicle, all scene information of traffic environment need to be obtained in time. The three-dimensional images of virtual reality are displayed in front of the control personnel, so that they can read and make remote control instructions quickly, and improve the remote-control performance of unmanned vehicles.

The remote control of the traditional unmanned vehicle mainly uses the camera to obtain the image. However, due to the limitations of the camera's visual field and the high latency and lack of data caused by environmental interference, remote control of unmanned vehicle becomes very difficult. Dąbrowska found that the vibration amplitude of the camera had a significant effect on the efficiency of the remote unmanned vehicle. At the same time, many research institutes apply lidar to the collection of scene images. Di Gennaro uses a multi-sensor method and remote sensing technique to obtain thermodynamic images through lidar mounted on unmanned aerial vehicles. Now, with the help of computer graphics and image processing technology, 3D visualization software has been widely used in many fields such as cultural relic protection, robot navigation and building manufacture [2]. It can be divided into three categories: modeling, platform and application. The modeling part is the core and foundation of the technology. For mobile robot remote control experiment, Kelly proposed to create a highly realistic 3D scene model based on virtual reality technology. First, a 3D terrain model is constructed by using 3D visualization technology based on virtual reality. Then, based on the registration images and radar data, realistic terrain rendering is implemented. Finally, the 3D terrain remote display system of unmanned vehicle remote control is designed and implemented, which has positive significance for the actual application of unmanned vehicles in the military field [3].

2. Experimental procedure

2.1. Virtual reality and 3D visualization

Virtual reality (VR) is a five-dimensional simulation of the real world, that is to say, besides the simulation of one dimension and three- dimensional space, it also includes the simulation of natural interaction. It is generated by a computer. By means of sensory modalities such as vision, hearing and touch, the system generates immersive interactive scene simulation. It's a computer system that creates and experiences the virtual world [4]. Virtual reality technology is a product of computer networks, graphics, artificial intelligence, sensors, information processing and other technologies. As shown in figure 1, it has three basic characteristics of immersion, interaction, and imagination, that is, the three "I". Immersion, also known as telepresence, means that in a virtual reality environment, users feel they become a "discoverer" and "actor"". At the same time, it can sense the reality of being a protagonist in a simulated environment. Interactivity refers to the degree to which the participants are able to operate within the virtual environment and receive feedback from the environment. That is, the user is the subject of interaction and is multi aware. Users can operate directly on objects in a simulated environment and get information or feel from the environment. These feelings can come from olfactory, gustatory, visual, and other sensory pathways [5]. Conceptual refers to the fact that users are immersed in multidimensional information spaces and acquire knowledge in a wide range of ways based on their perceptions and cognitive abilities. In the process of perception, they will exert their subjective initiative, get inspiration, seek answers, and form new concepts. In essence, the construction of virtual reality system is to create an information environment that enables participants to be in an immersive immersion, perfect interaction and inspiring ideas.

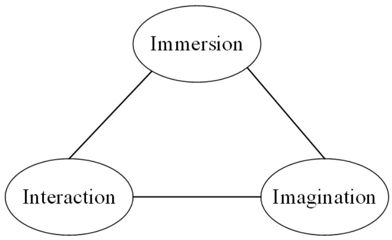

Fig. 1. Three "I" features of virtual reality

Three-dimensional visualization is the most important form of virtual reality technology. After a series of transformations, it converts the original analog data of the computer into the image that can be displayed. The idea is to convert abstract information into a format that can be understood by the human perception system. 3D terrain visualization is to build a 3D terrain model based on computer graphics and image processing technology. It is a reappearance of the real world's geographical environment with multi-levels and high fidelity. The realistic terrain model combines 3D spatial data with real image information It gives the image information obtained by the camera to the terrain surface in the form of texture mapping. At the same time, it can display complex terrain, geometric structure and surface properties, and satisfy people's demand for visual effects very well. A full realistic 3D modeling includes modeling objects, rich geometric information, and complete texture information. The specific process is shown in Fig. 2.

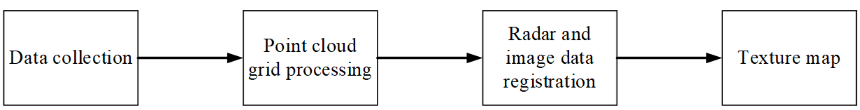

Fig. 2. 3D model building process

Firstly, lidar is used to obtain the point cloud data of the surface and depth information of the modeling object, and the corresponding image information is captured at the same time as the camera capture. In order to avoid the phenomenon of image information discarding, triangular meshes are used to process the point cloud data. After the lidar data are processed, the geometric model is obtained, and the camera image and the geometric model are corresponding to each other in space. Each point in the model is visible on the image, and the corresponding relation between the geometric model and the camera image is obtained [6]. After the registration, the corresponding coordinates of each vertex of the geometric model in the texture image can be determined. The texture map can be completed by drawing the engine to calculate the texture coordinates corresponding to the pixels in each mesh unit.

2.2. Texture mapping and image stitching

The camera projected the 3D scene in the real world onto the two-dimensional plane of the camera sensor, and completed the mapping from 3D space to two-dimensional plane. The accurate mapping relation is an important condition to assign correct texture information to the geometric surface of the modeling object in three-dimensional reconstruction. The image captured by the camera is converted into a digital image in a computer, and is represented by an array of $M \times N$ dimensions. Each element in the image is called a pixel, and the corresponding value is the brightness or RGB value of the image point. The coordinates of each pixel are represented by two values (U, V), representing the number of rows and columns of the pixel in the array of images, respectively. The origin O_1 is defined at the intersection of the camera's optical axis and the image plane, and the X and Y axes are parallel to the U and V axes, respectively. The coordinates of O_1 in the U and V coordinate systems are set to $(u_0,\ v_0)$. The physical dimensions of each pixel in the X direction and the Y direction are dx and dy. The coordinates of any pixel in the two coordinate systems are as follows:

$$u = \frac{x}{dx} + u_0, \quad v = \frac{y}{dy} + v_0\,. \tag{1}$$

The internal and external parameters can be obtained by the calibration of the camera [7]. When the position of radar and camera is fixed, the calibration relation between lidar and camera can be determined by the process shown below, and the unique corresponding point of radar data point in the camera image can be obtained. The known $[xyz1]^{\mathrm{T}}$ coordinates are regarded as coordinates of a point in the radar coordinate system. They are the corresponding coordinate of the point in the image coordinate system, which can be seen by the direct calibration method of the lidar and the camera:

$$\begin{bmatrix} u & v & 1 \end{bmatrix}^{\mathrm{T}} = G \begin{bmatrix} x & y & z & 1 \end{bmatrix}^{\mathrm{T}}. \tag{2}$$

According to formula (2), the equation $G = K[R|T]M$ can be obtained. The transformation matrix G can be used to find the only pixels corresponding to the lidar data points in the image. The matrix K contains the intrinsic parameters of the camera. The rotation matrix R and the translation vector T are the external parameters of the camera, and the orientation and position of the camera are determined.

The general process of texture mapping is to associate the multiple deformation vertices with their texture coordinates in the texture space during the modeling phase. Through the calibration of camera and radar, the elevation map triangle mesh vertex is mapped to the newest image acquired by camera. Each LIDAR point has a unique pixel corresponding to the data [8]. At this point, the task of the

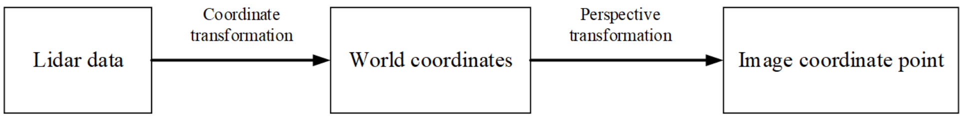

Fig. 3. Direct calibration principle of lidar and camera

rendering engine is to find out the corresponding texture coordinates for each pixel in each polygon. In the mainstream polygon rendering technology, the properties of the pixels inside the polygon are obtained by interpolating the properties of the vertices of the polygon. The triangular surface of the texture image is mapped to the target image one by one, so that the triangle in the source image is transformed into a triangle in the target image. Linear interpolation formula is shown as follows: $P_1(x_1, y_1)$, $P_2(x_2, y_2)$ and $P_3(x_3, y_3)$ are the three vertexes of a triangle. Symbols $P_a(x_a, y_k)$ and $P_b(x_b, y_k)$ are the two points over the P_k sweeping line

$$
\begin{aligned}
P_a &= \frac{1}{y_1 - y_2}\left[P_1(y_k - y_2) + P_2(y_1 - y_k)\right] , \\
P_b &= \frac{1}{y_1 - y_3}\left[P_1(y_k - y_3) + P_3(y_1 - y_k)\right] , \\
P_k &= \frac{1}{x_b - x_a}\left[P_a(x_b - x_k) + P_b(x_k - x_a)\right] . \qquad (3)
\end{aligned}
$$

The process of image mosaic is to transform images with overlapping to the same coordinate system, then the greater picture was synthesized. Image acquisition takes into account the focal length, type, position and motion state of the camera. Image preprocessing is the basic operation of digital image filtering, distortion correction, histogram processing and so on. The image matching process is to align the multiple images acquired at different times, locations and different cameras in space. Image transformation is a model transformation of the established image based on the matching point. The image to be spliced is converted to a unified reference image coordinate system. The process of image fusion is to eliminate the stitching seam and the matching error in the coincident area of the mosaic image.

2.3. Terrain geometry modeling and rendering methods

In order to make 3D realistic scene model meet the remote control requirement of unmanned vehicle, we must pay attention to the efficiency of data processing and the effect of 3D rendering of scene. In order to ensure the modeling effect, we need to preprocess the original cloud data, such as filtering, denoising, data segmentation and so on, before we build the 3D terrain model based on lidar data. In order to achieve the effect of photo realistic terrain visualization, fast terrain rendering, background region rendering and terrain visualization modeling are carried out on the basis of geometric model and image data obtained from registration.

For the 3D terrain construction method based on lidar data, the interpolation technique is first used to obtain continuous and smooth point cloud data after preprocessing the original point cloud data. Then the triangulation algorithm is used to obtain the terrain surface. The terrain mesh obtained by interpolation and subdivision can only represent the geometry information of the far end scene. Texture information is also a necessary part of the remote scene model to achieve realistic. The camera image is simultaneously acquired by a plurality of cameras loaded by an unmanned vehicle [9]. These images are used for stitching, and the stitching results are applied to Billboard vision. This can show greater vision space, realize the realistic sense of the single frame terrain visualization.

2.4. Three-dimensional remote display system of unmanned vehicle

A 3D terrain remote display system for unmanned vehicle remote control is designed and implemented. The Basler SC A1400-17gm model camera and Velodyne HDL-64E S2 64 line lidar are used as sensors. The sensor is fixed on an unmanned vehicle platform, and simultaneously generates 3D point data and corresponding image data in the radar coordinate system. The data is transmitted to the remote control platform by wireless link, and the wireless network is used to transmit the information. The virtual simulation environment is used as visual feedback, so that the operator can switch the viewpoint from the driving position in the reconstructed scene and the position near the unmanned vehicle. This process realizes the observation of the environment from the virtual viewpoint. In the integrated development environment, a series of terrain and ground modeling experiments are carried out by using the programming program, and the 3D visualization of the lidar image and the camera image data is realized. A visualized 3D terrain model of unmanned vehicle remote control is established, and the validity and feasibility of the system are verified.

3. Results and discussion

3.1. A telepresence scheme for realistic scenes

This paper starts with two aspects of modeling realism and high efficiency. The high precision 3D data of the measured scene is acquired by lidar, and the 3D triangle mesh model of the scene is rapidly built. Through the calibration of radar and camera, combined with the image information or coloring scheme of camera, a 3D realistic scene model with rich surface information and remote control operation can be reconstructed. The data processing flow of the design scheme. The overall design scheme includes two approaches, common cloud point data preprocessing and 3D terrain geometry modeling. The realistic terrain rendering method based on projection texture mapping and the fast terrain rendering method based on point coloring are implemented. The remote scene expresses the Billboard part and the ground surface model representation of non ground points.

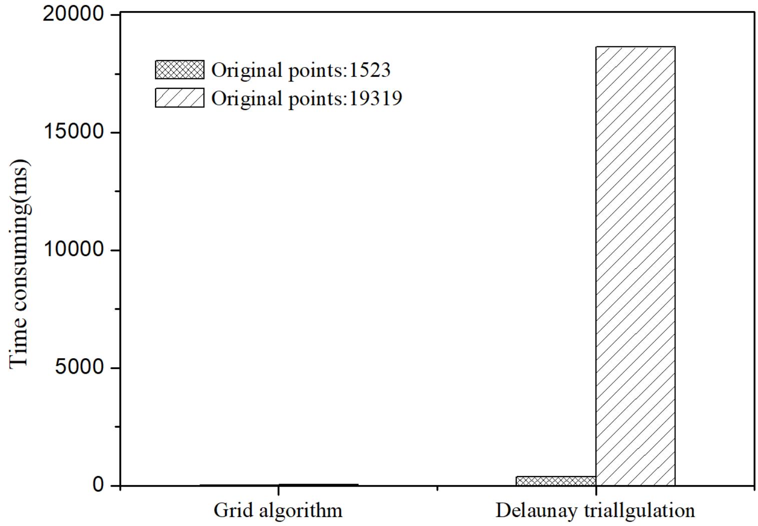

Fig. 4. Time consumption contrast between two algorithms

3.2. Three-dimensional terrain rendering results

The 3D mesh surface model of the system is constructed by elevation map with a uniform square grid. However, the traditional Delaunay algorithm combines the similar points in the original discrete point set data according to certain criteria to form the optimal triangle. In order to make the point centralized, each data point becomes a vertex of the triangular mesh, and the original scattered point sets are used to generate continuous irregular triangulated networks that can cover the whole point set region. The time-consuming comparison results are shown in Fig. 4. The Delaunay algorithm can produce better segmentation effect, but it takes too long. In order to improve the computational efficiency of terrain subdivision and make it applicable to the unmanned vehicle simulation environment with high real-time requirement, the terrain grid is verified by experiments. The algorithm is simple and time-consuming. And the segmentation effect can approximately express the relief of the terrain, satisfy the demand of mapping and remote control, and have the whole optimum property.

Texture information can not only give a realistic visual effect, but also make up for the lack of geometric modeling accuracy. Therefore, taking eight-line lidar data and camera image data as an example, the texture information obtained by using some technical means is added to the camera. This verifies the validity of the rendering. Texture mapping and terrain visualization for eight-line radar data are shown in Table 1.

Compared with traditional video remote manipulation images, the texture of realistic terrain model after texture mapping has abundant texture information, and can realize the observation of environment from different viewpoints [10]. With the help of the terrain model of 3D terrain and image information, the geometry information of the scene can be enriched, and the exit path and the accessible road can be easily judged. As shown in Table 1, interpolation time and texture coordinate computation time are increased as the number of meshes increases, while rendering time is almost zero. In the texture mapping of 400×400 mm mesh size, good visual effects have been achieved and the time consuming is about 55 ms/frames.

Table 1. Experimental parameters

Line No.	Mesh size (mm)	Mesh number	Interpolation time (ms)	Texture coordinate calculation time (ms)	Render time (ms)
5292	400×400	25×45	16	31	8
10260	300×300	34×60	15	47	14
14688	250×250	40×71	32	78	16

Compared with texture mapping, point coloring method avoids the process of texture interpolation, and the rendering efficiency is much higher than that of texture mapping method. However, the resolution of radar data is far below the image resolution. A large amount of image information is discarded if it is directly colored by the point cloud. This effect cannot meet the remote-control requirements. Therefore, the patch blocks are used instead of point coloring points to fill the gap of the point cloud, so as to compensate for the loss of the information when the resolution of the radar is lower than the resolution of the image. Under the condition that the number of interpolation points is 5292, the mesh size and the number are 400×400 mm and 25×45 mm, respectively, the two methods of terrain rendering based on projection texture mapping and point based shading are compared. The result is shown in Fig. 5. When using the same size mesh, the interpolation time of the two rendering methods is different from that of the texture/color coordinate. But for the rendering time, it is more efficient to use the texture mapping based on realistic terrain modeling method.

In order to show the special effects of 3D scene realistically, billboard method is used to display the background area. The billboard is the projection of the camera image, the surface must be parallel to the camera plane, and located on the camera's optical axis. Its size covers the projection area, and the image moves with the motion of the vehicle. When the virtual viewpoint is the same as the camera's actual location, the billboard is barely perceptible. As the virtual viewpoint moves farther away from the camera, the deformation becomes apparent. Billboard performs better when viewed from a distant point of view, but when the point of view approaches, its effect becomes dramatically worse. In the process of image matching, NCC operator is used as the similarity measure operator to match the two images. Image mosaics are performed using an affine transformation model with the parameter 6. Finally, in order to eliminate the stitching difference between two images, a smooth and natural stitching result is obtained. The pixel values of the pixels on the two sides of the

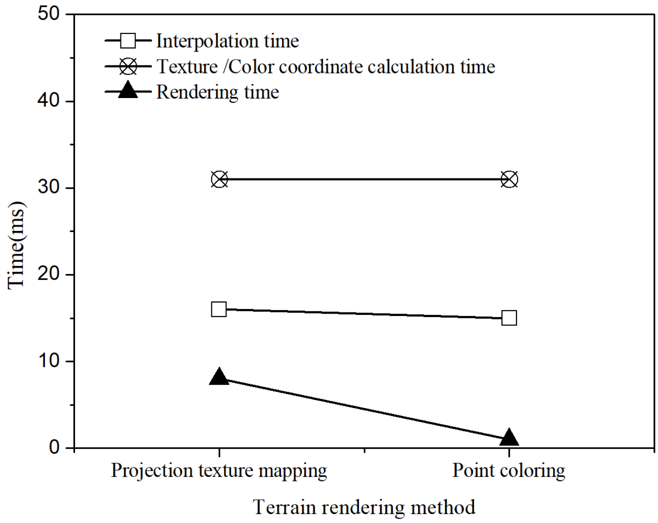

Fig. 5. Time consumption contrast between two rendering methods

joint and the distance between the pixels and the joint are weighted by the method of linear fade. The pixels of the overlapped region are mixed to obtain the final image results. Integrating the terrain model with the object model, the complete remote scene expression is obtained.

3.3. Design and implementation of 3D remote display system for unmanned vehicle

Virtual reality uses digital graphics models to recreate scenes in the real world. The effects and the fidelity of the scene we see are related to the model. The basic hardware settings at the system level include the remote control station, the on-board unit equipped with the lidar, the camera, and the wireless communication module for transmitting instructions and video. In the Visual C++ integrated development environment, the application of Open Scene Graph is applied to carry out a series of experiments on terrain and ground features modeling. The 3D visualization of imaging lidar images and camera images is realized, and a 3D visualization model of unmanned vehicle remote control is established.

Figures 6 and 7 are time consumption comparison of terrain modeling between real terrain rendering module and point shading terrain rendering module. The original LIDAR point cloud number is about 11966, and the interpolation points are 11484 and 44814 respectively. The corresponding mesh sizes and numbers are 400×400 mm and 91×32 mm, 200×200 mm and 194×87 mm, respectively.

Using 400×400 mm mesh size to interpolate texture mapping, the time consumed during modeling is approximately 141 ms/frames. Texture mapping using 200×200 mm interpolation is similar to it. But the time consumption of interpolation and texture coordinate calculation is obviously increased. In addition, the

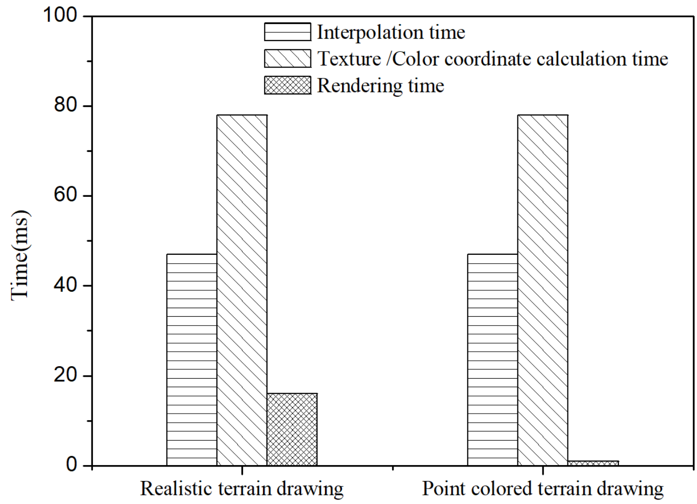

Fig. 6. Time consumption comparison results (mesh size 400×400)

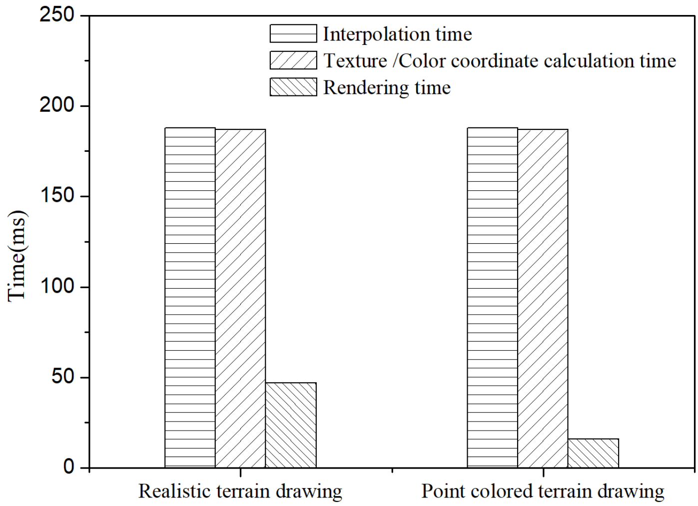

Fig. 7. Time consumption comparison results (mesh size 200×200)

400×400 mm mesh size interpolation of point cloud coloring models has large gaps and is inconvenient for operator driving observation. After shrinking the mesh to 200×200 mm, the coloring effect basically meets the needs of remote display, but the time consumption reaches 391 ms/frames. The computation of texture coordinates of interpolation and point clouds takes a great deal of time, while the time consumption of rendering is significantly lower than that of realistic terrain rendering. The modeling results of the two functional modules can meet the remote control requirements of unmanned vehicle remote control. In contrast, photorealistic rendering module works better and is more realistic. After fine interpolation, the point shading terrain rendering module can achieve the approximate realistic shading effect. But the time consumption of interpolation processing is relatively large. In time processing, the two methods have greater room for improvement.

4. Conclusion

The remote visualization system of unmanned vehicle remote control is used as an application background, and the 3D visualization terrain modeling method is studied from two aspects of theory and practice Firstly, the basic theories and algorithms involved in the data fusion techniques of lidar and camera images are analyzed. On this basis, a rapid terrain geometry modeling scheme is proposed. Considering the two aspects of modeling effect and rendering efficiency, realistic rendering based on projection texture mapping is applied to realize realistic rendering of terrain model with rich scene information. Then, the point rendering terrain visualization method with adjustable resolution is used to improve the rendering efficiency of the model. At the same time, the point size control function makes the interpolation time consumption and the modeling effect reach a good compromise. Experiments show that the two methods can realize the observation of the virtual viewpoint environment without the sensor data angle. High precision 3D data of measured terrain are acquired by lidar. A 3D terrain model with abundant surface information is reconstructed by combining the image information of the camera or the coloring scheme. The system can support scene roaming at any point of view, which overcomes the restriction of the limited observation field of vision in the traditional vehicle remote control mode, and meets the demand of unmanned vehicle remote display.

The system only realized the terrain modeling of unmanned vehicle remote control, and did not study the modeling of non-surface objects. At the same time, the system does not have the Billboard vision display function based on image mosaic. Based on the visualization of single frame terrain, the future research can provide location and status information by means of navigation system, and integrate and map multi frame radar data. In this way, more abundant information can be expressed, and a realistic offline 3D scene map can be set up, which can better meet and meet the needs of unmanned vehicle applications.

References

[1] A. J. Mathews, J. L. R. Jensen: *Visualizing and quantifying vineyard canopy LAI using an unmanned aerial vehicle (UAV) collected high density structure from motion point cloud.* Remote sensing *5* (2013), No. 5, 2164–2183.

[2] A. Dąbrowska, M. B. Jaskółowski, A. Rubiec: *Cameras vibrations influence on efficiency of teleoperated unmanned ground vehicle.* IEEE International Conference on Methods and Models in Automation and Robotics (MMAR), 29 August–1 September 2016, Miedzyzdroje, Poland, IEEE Conference Publications, (2016), 772–777.

[3] S. F. Di Gennaro, A. Matese, B. Gioli, P. Toscano, A. Zaldei, A. Palliotti,

L. Genesio: *Multisensor approach to assess vineyard thermal dynamics combining high-resolution unmanned aerial vehicle (UAV) remote sensing and wireless sensor network (WSN) proximal sensing.* Scientia Horticulturae *221* (2017), 83–88.

[4] A. Kelly, E. Capstick, D. Huber, H. Herman, P. Rander, R. Warner: *Real-time photorealistic virtualized reality interface for remote mobile robot control.* Robotics Research, Part: Springer Tracts in Advanced Robotics, Book series STAR *70* (2011), 211–226.

[5] G. M. E. Sanches, T. Van Renterghem, K. Sun, B. De Coensel, D. Botteldooren: *Using virtual reality for assessing the role of noise in the audio-visual design of an urban public space.* Landscape and Urban Planning *167* (2017), 98–107.

[6] M. Kulawiak, M. Kulawiak: *Application of web-GIS for dissemination and 3D visualization of large-volume LiDAR data.* The Rise of Big Spatial Data, Part: Lecture Notes in Geoinformation and Cartography, Springer Nature (2017) 1–12.

[7] S. Iwasaki, T. Narumi, T. Tanikawa, M. Hirose: *Guidance method to allow a user free exploration with a photorealistic view in 3D reconstructed virtual environments.* Proc. International Conference on Distributed, Ambient, and Pervasive Interactions (DAPI), 9–14 July 2017, Vancouver, BC, Canada, Conference proceedings DAPI, Part: Lecture Notes in Computer Science, Springer Nature *10291* (2017), 347–357.

[8] J. Liu, J. Li, W. Xu, Y. Liu, M. Zhang: *Mosaicing of fisheye images through conformal projection and sphere matching.* Proc. International Conference on Graphic and Image Processing (ICGIP), 29 October 2016, Tokyo, Japan, Processing SPIE *10225* (2016).

[9] Y. Y. Nie, Z. Tang, J. F. Yu, Y. R. Zhu, J. Chang, J. J. Zhang, S. H. Guo, Y. Su: *Image-based 3D scene reconstruction and rescue simulation framework for railway accidents.* Proc. IEEE International Conference on Virtual Reality and Visualization (ICVRV), 24–26 September 2016, Hangzhou, China, IEEE Conference Publications, (2016), 335–340.

[10] G. Zhong, H. Deng, Y. Kobayashi, H. Wang: *Theoretical and experimental study on remote dynamic balance control for a suspended wheeled mobile manipulator.* Nonlinear Dynamics *79*, (2015), No. 2, 851–864.

Received July 12, 2017