한국정보전자통신기술학회논문지(jkiiect)'19-12 Vol.12 No.6
논문 19-12-06-605 http://dx.doi.org/10.17661/jkiiect.2019.12.6.605

# High efficient 3D vision system using simplification of stereo image rectification structure

Sang Hyun Kim*

# 스테레오 영상 교정 구조의 간략화를 이용한 고효율 3D 비젼시스템

김상현

**Abstract** 3D Vision system has many applications recently but popularization have many problems that need to be overcome. Volumetric display may process a amount of visual data and design the high efficient vision system for display. In case of stereo system for volumetric display, disparity vectors from the stereoscopic sequences and residual images with the reference images has been transmitted, and the reconstructed stereoscopic sequences have been displayed at the receiver. So central issue for the design of efficient volumetric vision system lies in selecting an appropriate stereo matching and robust vision system. In this paper, we propose high efficient vision system with the reduction of rectification error which can perform the 3D data extraction efficiently with low computational complexity. In experimental results with proposed vision system, the proposed method can perform the 3D data extraction efficiently with reducing rectification error and low computational complexity

**Key Words :** 3D display, 3D Vision, Image rectification, Stereo image, Video coding

## 1. Introduction

Recently, popularity of 3D video has been growing significantly and it may turn into a home user mass market in the near future. An emerging 3D video formats and standards are given, which are mostly related to specific types of applications and 3D displays [1]. This includes conventional stereo video, multiview video, video plus depth, multiview video plus depth and so on.

3D video is commonly understood as a type of visual media that provides depth perception of the observed scenery. It is also referred to as stereo video. Such 3D depth perception can be provided by 3D display systems which ensure that the user sees a specific different view with each eye. The stereo pair of views must correspond to the human eye positions. Then the brain can compute the 3D depth perception. Most 3D display systems require wearing specific glasses to ensure separation of left and right view which are displayed simultaneously.

Awareness of and interest in 3D video is rapidly increasing, among users who wish to

*Corresponding Author : School of convergence and fusion system engineering, College of science and technology, Kyungpook National University (shk@knu.ac.kr)

Received December 02, 2019 Revised December 12, 2019 Accepted December 12, 2019

experience the extended visual sensation, as well as among content producers, equipment providers, and distributors. At the same time technology is maturating from capture to display. The market of 3D cinema is expected to continue growing rapidly over the next years. More and more cinemas are being equipped with 3D technology. With the content being produced, 3D video is also an increasingly interesting technology for home user living room applications. 3D video content will arrive to the home by 3D DVD, Internet, 3DTV broadcast, etc. Currently, there is a great variety of different 3D display systems designed for the home user applications, starting from classical 2-view stereo systems with glasses. More sophisticated candidates for 3D vision in living rooms are multiview auto-stereoscopic displays, which do not require glasses. They emit more than one view at a time but the technology ensures that users only see a stereo pair from a specific viewpoint. Today's 3D displays are capable of showing 9 or more different images at the same time, of which only a stereo pair is visible from a specific viewpoint. This supports multi-user 3D vision without glasses in a living room environment. Motion parallax viewing can be supported if consecutive views are stereo

There are a lot of different 3D video formats available and under investigation. They include different types of data, mostly related to specific types of displays. A variety of compression and coding algorithms are available for the different 3D video formats. Some of these standards and coding algorithms are standardized by MPEG, since standard formats and efficient compression are crucial for the success of 3D video applications [2]. A generic, flexible and efficient 3D video format that would serve a wide range of different 3D video systems is highly desirable in this context. Therefore MPEG is currently investigating such a new generic 3D video standard. For efficient 3D coding, the accurate disparity map is essential. In this paper, we propose high efficient vision system using simplified image acquisition sytem. The stereo video coding for 3D broadcasting is presented in section 2. The proposed stereo image system for 3D display is described in section 3, and experimental results are shown in Section 4. Finally conclusions are given in Section 5.

## 2. STEREO VIDEO CODING FOR 3D VISION SYSTEM

The 3D video representation which captures by 2 or more cameras and video signals may have undergone some processing steps like rectification. The video signals are meant in principle to be directly displayed using a 3D display system, though some video processing might also be involved before display.

Compared to the other 3D video formats the algorithms are the least complex. It can be as simple as only separately encoding and decoding the multiple video signals. Only the amount of data is increasing compared to 2D video. Coding efficiency can be increased by combined temporal and interview prediction.

MPEG-2 provided a corresponding standard such as MPEG-2 Multiview

Profile. Recently, Stereo SEI (Supplemental Enhancement Information) message was added to the latest and most efficient video coding standard H.264/AVC, which implements inter-view prediction. For more than 2 views this is easily extended to Multiview Video Coding (MVC). A corresponding MPEG-ITU standard is an extension of H.264/AVC [3]. It can also be applied to 2 views. MVC is currently the most efficient way for stereo and multiview video coding, whereby the performance of a solution based on the H.264/AVC Stereo SEI message is similar for the stereo case. A simple way to use existing video codecs for stereo video transmission is to apply temporal or spatial interleaving. A problem is that there is no corresponding standard available. There is no way to signal the use of interleaving to the decoder. A normal video decoder would decode the stereo video incorrectly.

An approach for efficient stereo video coding is derived from the binocular suppression theory. If one of the images of a stereo pair is low-pass filtered, the perceived overall quality of the stereo video will be dominated by the higher quality image. The perceived quality will be as if both images are not low-passed. Based on that effect, mixed resolution stereo video coding can be derived. Instead of coding the right image in full resolution it is down sampled to half or quarter resolution. The baseline is fixed from capturing. Depth perception cannot be adjusted to different display types and sizes. The number of output views can not be varied.

Broadcasting based on home servers will enable wide range of viewing styles with the use of meta data provided by broadcasters. For instance, viewers will be able to watch any program at their convenience and retrieve only the information of interest. The advent of the home server will be the beginning of full scale integrated services television, which will integrate various new services exploiting the home server's broadcasting, communications, and storage functions. The linked services provided over home servers will require a safe distribution environment for program content and metadata. These will be able to acquire the licensing information necessary for content viewing via broadcasting and communications and control content utilization.

Stereoscopic and multiview images are still considered as the most adaptable methods to the current plane imaging systems, among the three dimensional imaging methods. The geometry for 3D data acquisition with stereo camera is shown in Fig. 1. As shown in Fig. 1, the stereo images obtained from left and right camera can extract the 3D data using depth map that represents the distance from camera. The full volume for the 3D image displays is still some distance away. This is not because of the lack of 3D imaging contents but because the quality of 3D images is usually somewhat worse than that of plane images, which is now in the super High Definition (HD) level.

The resolution of super HD is four times higher than that of Full HD, which is considered as having a remarkable image quality. This is why major researches are interested in stereoscopic imaging with full display panel resolution.

The recent development of high speed LCD which is capable of operating 240Hz frame rate makes feasible the stereoscopic imaging. This LCD is used either as display panel or parallax barrier [4]. In the display panel, left and right eye images are switched time-sequentially to display stereoscopic images with each view image having full display panel resolution. The time sequential display method is especially advantageous to the displays with a very limit resolution, as in mobile phones. The time sequential method allows making CIF image resolution for each view image in the mobile displays.

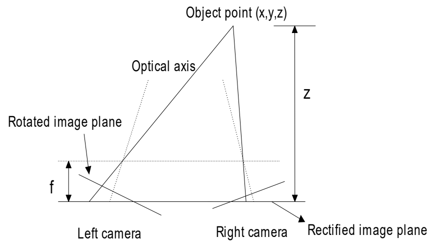

Fig. 1. Stereo Imaging system for 3D data acquisition

Compared with the stereoscopic image, 3D images based on the multiview images have been considered as providing more natural depth sense and comfort to viewers. But the resolution of each view image is low compared with the panel resolution and becomes lower as the number of different view images to be displayed increases. Using more projectors, the resolution of each view image is not even further increased but also the number of multiview images. The cost of building them will be very high and aligning the projectors requires much more efforts than other multiview 3-D imaging methods. These problems will be aggravated more as the number of the projectors increases.

In the electro-holographic imaging areas, only few new developments related with fast calculation algorithm for Computer Generated Hologram have been developed. To make fit the flat panel displays to 3D has a significant meaning because 3D displays are getting out from the subordinate relationship to flat panel displays. Recording and displaying 3D natural scene based on fringe pattern calculation with the image depth information obtained from a depth camera or a multiview camera array represent accurate result.

The multi-view video has advantages of providing multiple viewpoints to users by capturing a three-dimensional scene from different camera positions. The 3DTV and free viewpoint TV have been discussed as main applications of the multi-view video, which need multi-view images and the corresponding depth maps to provide 3D videos or the multi-view video as inputs for these applications. Stereo image have horizontal disparities and vertical mismatches between neighboring views. The stereo system as shown in Fig. 1 can generate depth maps from images having horizontal disparities with stereo image rectification [5]. If multi-view images have not vertical mismatches but disparities only in the horizontal direction, these can generate high quality depth maps much faster than the case of having vertical mismatches. For 3D broadcasting, the accurate depth map using stereo image system is essential. The proposed stereo system employs the simple rectification structure which can be performed the real-time 3D broadcasting efficiently.

## 3. PROPOSED STEREO IMAGE SYSTEM FOR 3D DISPLAY

The 3D data can be extracted by stereo matching with image rectification. Image rectification is a kind of image transformation. Stereo image rectification is a process that makes epipolar lines of two images captured at different positions parallel each other. Then, vertical coordinates of all image points of two images become equal, and there remain only horizontal disparities as shown in Fig. 2.

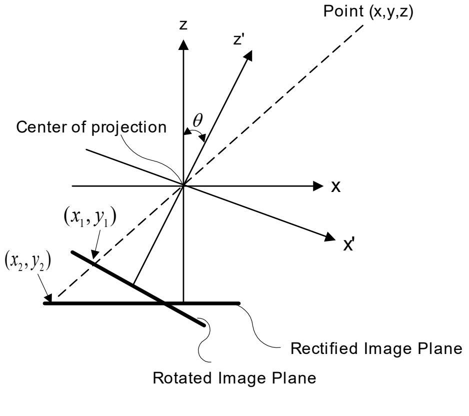

Fig. 2. Simplified geometry for stereo image rectification

As shown in Fig. 2, the geometry for stereo image rectification has been simplified by the correspondence of projection center. The geometrical relationship between rotated and rectified image plane, and where $A$ and $R$ represent perspective projection and rotation matrix, respectively.

$$x = R x' \quad x_1 = A x' \quad x_2 = A x \qquad (1)$$

$$x_2 = AR x' = AR A^{-1} x_1 \qquad (2)$$

$$A = \begin{pmatrix} fS_x & 0 & C_x \\ 0 & fS_x & C_y \\ 0 & 0 & 1 \end{pmatrix} \qquad (3)$$

$$R = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \qquad (4)$$

The rectification can be performed from $(x_1, y_1)$ to $(x_2, y_2)$ plane to plane by equation

$$x_2 = f\frac{\cos\Theta * x_1 + \sin\Theta * f}{-\sin\Theta * x_1 + \cos\Theta * f}$$

$$y_2 = f\frac{y_1}{-\sin\Theta * x_1 + \cos\Theta * f} \qquad (5)$$

The line through two camera centers becomes the baseline. Then, we can get rectified images by applying the simplified rectification transformation to stereo images.

## 4. Experimental results

Table 1 shows the rectified control points in stereo images with simplified rectification process. As shown in Table 1, the rectified control points have a place on parallel epipolar lines which can reduce the computational complexity to extract the disparity data. The stereo matching with simplified epipolar lines can perform the encoding process for 3D broadcasting efficiently with low complexity.

Table 1. Rectified control points for 3D data acquisition.

| Control points using Conventional Method | | Control Points using Proposed Method | |
|---|---|---|---|
| Vertical Control Points | Plane Error (pixel) | Rectified Points | Plane Error (pixel) |
| 341.9 / 328.9 | 13.0 | 353.1 / 353.3 | 0.2 |
| 346.6 / 333.5 | 13.1 | 362.4 / 362.6 | 0.2 |
| 351.2 / 338.1 | 13.1 | 371.8 / 371.9 | 0.1 |
| 355.6 / 342.5 | 13.1 | 381.1 / 381.0 | 0.1 |
| 359.7 / 346.5 | 13.2 | 390.0 / 390.0 | 0.0 |
| 363.9 / 350.7 | 13.2 | 399.2 / 399.2 | 0.0 |
| 368.1 / 354.8 | 13.3 | 408.5 / 408.5 | 0.0 |
| 403.1 / 386.4 | 16.7 | 425.4 / 425.6 | 0.2 |
| 406.7 / 390.3 | 16.4 | 434.8 / 434.7 | 0.1 |
| 410.5 / 394.0 | 16.5 | 443.9 / 443.9 | 0.0 |

In table 1, vertical control points represent conventional control points which show points to extract disparity, and the plane error represents the vertical difference between vertical points. The reduction of the plane error is key element to reduce stereo matching error and the computational complexity. When the plane error is large and vertical search region is N rows, the computational complexity in conventional method increases N times compare with proposed method reducing plane error. In case of high efficient vision system, reduction of plane error is essential. So the proposed vision system with the reduction of rectification error can perform the 3D data extraction efficiently with low computational complexity.

## 5. CONCLUSIONS

The multiview 3D images need a very high resolution display panel for multiview image display. The guidelines for measuring the quality of 3D images and standard methods of measuring factors affecting the quality should be developed with new flat panel displays based on holography and super-multiview. Depth resolution of 3D imaging will become a major factor for determining the quality because users can interact with 3D scenes on the displays. In this paper, we have presented stereo imaging system with simplified rectification to extract 3D data efficiently.

The simplified rectification process can reduce rectification error with low computational complexity and can be applied to multi-view video applications with 3D vision system and the volumetric medical imaging system.

## REFERENCES

[1] Q. Chang and T. Maruyama, "Real-time stereo vision system: A multi-block matching on GPU" *IEEE Access*, Vol. 6, pp. 42030-42046, 2018.

[2] F. Zhong, and C, Quan, "A single color camera stereo vision," *IEEE Sensors*, vol. 18, no. 4, pp. 1474-1482, Dec. 2017.

[3] W. Wang, J. Yan, N. Xu, Y. Wang, and F.-H. Hsu, "Realtime high quality stereo vision system in FPGA," *IEEE Trans. Circuits and Systems for Video Technology*,

vol. 25, no. 10, pp. 1696-1708, Oct. 2015

[4] Y. Wang, L. Q. Hao, X. Wang, D. L. Lau, and L. G. Hassebrook, "Robust active stereo vision using Kullback-Leibler Divergence," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 34, no. 3, pp. 548-563, Mar. 2013.

[5] L. Sawalha, M. P. Tull, M. B. Gately, J. J. Sluss, M. Yeary, and R. D. Barnes, "A large 3D swept-volume video display," *Journal of Display Technology*, vol. 8, no. 5, pp. 256-268, May 2012.

## Author Biography

**Sang Hyun Kim** **[정회원]**

He received the B.S. and M.S. degrees in electronic and control engineering from Hankuk University of Foreign Studies, in 1997 and 1999, respectively, and the Ph.D. degree in electronic engineering from Sogang University in 2003. In 2003 and 2004, he worked on the Digital Media Research Laboratory in LG Electronics Inc., as a Senior Research Engineer. In 2004 and 2005, he also worked on the Computing Laboratory at Digital Research Center in Samsung Advanced Institute of Technology, as a Senior Research Member. In Mar. 2005 and Feb. 2015, he had been with the school of electrical engineering at Kyungpook National University as an associate professor. Since Mar. 2015, he has been with the school of convergence and fusion system engineering at Kyungpook National University as a professor.

〈Research Interests〉 computer vision, video coding, and convergence system