POSTER: iGLANCE project: free-viewpoint 3D video

Svitlana Zinger
Eindhoven University of Technology
P.O. Box 513
5600MB Eindhoven, Netherlands
s.zinger@tue.nl

Daniel Ruijters
Philips Healthcare
Veenpluis 6
5680DA Best, Netherlands
danny.ruijters@philips.com

Peter H. N. de With
Eindhoven University of Technology
P.O. Box 513
5600MB Eindhoven, Netherlands
P.H.N.de.With@tue.nl

ABSTRACT

The iGLANCE project aims at researching the methods of receiving and rendering of free-view 3D-TV sequences. We describe the ongoing iGLANCE project and introduce the iGLANCE decoding and rendering system. It consists of free viewpoint video interpolation and coding. One of the key components of this system is view interpolation and its related challenges. There are two main application fields within iGLANCE: consumer and medical systems. In this article we discuss an example use case for a medical application that illustrates the need for the research and development in the iGLANCE framework.

Keywords

3D video, view interpolation, autostereoscopic display, free-viewpoint.

1. INTRODUCTION

Multi-view autostereoscopic displays allow depth perception of displayed scenes without the aid of additional goggles. The stereoscopic effect is obtained by presenting a slightly different image of a 3D scene to the left and the right eye of a viewer (see Figure 1). Since a multi-view display emits more than two views, a viewer can be positioned anywhere within the stereoscopic range, and multiple viewers can perceive depth at the same display.

The combination of autostereoscopic displays with free-viewpoint video creates opportunities with respect to interaction between the user and the broadcasted content. Furthermore, the view interpolation techniques allow reducing the amount of information that has to be coded and transmitted. To achieve these goals, new algorithms for image and video processing are required.

The iGLANCE project has started in October 2008 and aims at investigating both the algorithmic solutions as well as the hardware design for the objectives described above. The project challenge originates from the use of multiple views of the same scene (multiple cameras) and the generation of arbitrary in-between views. This involves advanced coding, compression and rendering of 3D video in HD resolution.

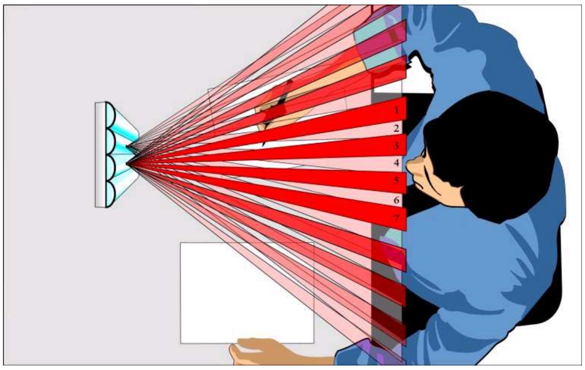


Figure 1. The multi-view autostereoscopic display.

The generating process of free-viewpoint 3D video, which is a research topic in computer vision, enables the user to view a scene from a freely chosen position. For example, a possible use case of iGLANCE is watching a concert or football game from the preferred viewpoint. The possibility to interactively choose a viewpoint in-between the broadcasted 3D video views distinguishes this projects from other research projects [Red02, Onu06] and present commercial 3D-TV solutions.

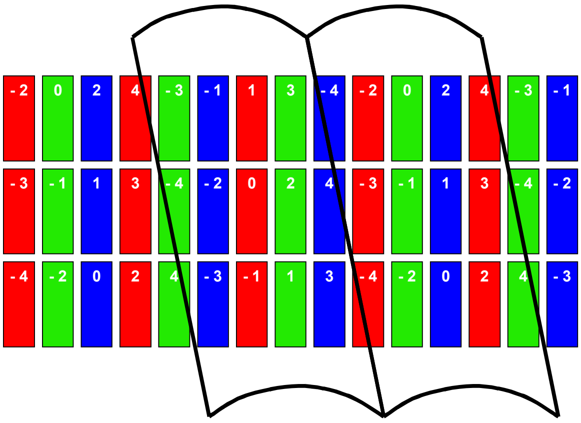
3D Video processing is emerging for both medical purposes and consumer electronics. In the medical domain, 3D data is generated by several imaging modalities, and after processing, visualization of this data on an autostereoscopic screen with free-viewpoint option will help to optimally navigate instruments during interventions. This type of data display is especially useful in minimally invasive interventional procedures.

2. MULTI-VIEW AUTOSTEREOSCOPIC DISPLAY

A multi-view autostereoscopic lenticular display consists of a cover sheet of cylindrical lenses (lenticulars) placed on top of an LCD, in such a way that the LCD image plane is located at the focal plane of the lenses [Ber99, Rui06]. The effect of this arrangement is that LCD pixels located at different positions underneath the lenticulars fill the lenses when viewed from different directions. Provided that these pixels are loaded with suitable stereo information, a 3D stereo effect is obtained, in which the left and right eye see different, but corresponding, information. A commercially available display offers nine distinct views, but our technology will be applicable to any number of views.

The fact that the different LCD pixels are assigned to different views (spatial multiplex), leads to a lower effective resolution per view than the intrinsic resolution of the LCD grid [Dod97]. In order to distribute this reduction of resolution over the horizontal and vertical axis, the lenticular cylindrical lenses are not placed vertically and parallel to the LCD column, but slanted at a small angle.

The resulting assignment of a set of LCD pixels is illustrated in Figure 2. Note that the red, green and blue color channels of a single pixel are depicted in different views, and that 9 different views are covered within a column of 3 pixels.



-2	0	2	4	-3	-1	1	3	-4	-2	0	2	4	-3	-1
-3	-1	1	3	-4	-2	0	2	4	-3	-1	1	3	-4	-2
-4	-2	0	2	4	-3	-1	1	3	-4	-2	0	2	4	-3

Figure 2. The cylindrical lenses depict every sub-pixel in a different view. The numbers in the sub-pixels indicate in which view they are visible.

The frustums that result from the different focal spot positions, are illustrated in Figure 3. The viewing directions of the frustums are not parallel to the normal of the screen, except for the central one. Therefore the corresponding frustums are asymmetric [Mau05]. The viewpoint coordinates of such parallel axis, asymmetric frustum perspective projections are determined by

$$\left(\frac{(x - n \cdot d) \cdot f}{f - z} + n \cdot d \ , \ \frac{y \cdot f}{f - z} \right), \qquad (1)$$

where f denotes the focal distance, n the view number and d the distance between the view cameras.

Figure 4 illustrates the images that result from rendering the scene from focal spot positions with an offset to the center of the screen.

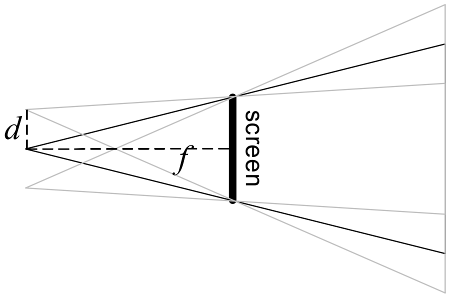


Figure 3. The frustums resulting from three different viewpoints.

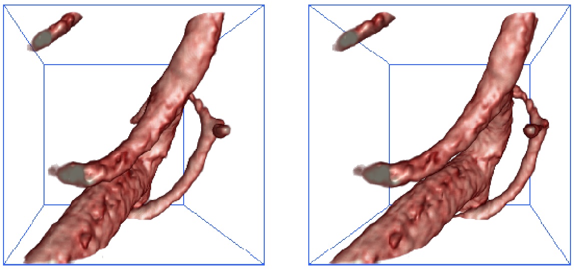


Figure 4. The same scene rendered from the most left and most right viewpoint.

3. FREE VIEWPOINT RENDERING

A key algorithmic development within iGLANCE concentrates on the creation of intermediate 3D video views, given several predetermined views of the same 3D scene. These algorithms include multi-view image interpolation, view warping and depth estimation [Mor08]. The input views are supposed to encode the conventional image (texture) and a depth or disparity map. Obtaining such depth images for real-world video sequences [Red02] is not an explicit objective of the iGLANCE project; the availability of such information is assumed as a precondition and can be obtained with off-line computations. For full

3D data, such as in virtual reality, computer games, or medical volumetric data, the generation of the depth information is more trivial.

The fundamental challenge that remains is displaying occluded parts of the scene [Vaz08]: in interpolated views, a part of the scene may become visible that has been hidden in any of the transmitted views, see Figure 5.

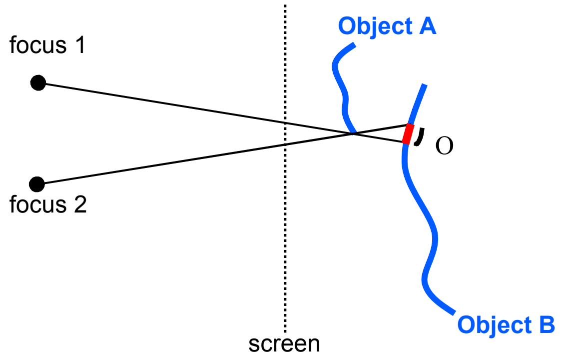


Figure 5. Illustration of occlusion: the area O of object B can only be seen from focus 2, thus is occluded from focus 1.

4. HARDWARE DESIGN

The hardware that is being developed for the iGLANCE project focuses on the reception and decoding of the multi-view video streams and performing the free-viewpoint interpolation.

Each video stream frame contains 2D texture and its respective depth map. Figure 6 defines the iGLANCE transmission chain, showing its main components.

Our research on view interpolation aims at finding the optimal methods for generating new views. The questions we are addressing include search for an optimal rendering method that produces good results in reasonable time, choice of the existing views (in time and space) for producing a new view, dealing with occlusions.

The outcome of this research will be beneficial for both consumer and medical applications. For the consumer applications our results will provide basis for free-viewpoint 3D-TV: the user will choose the viewpoint, and the iGLANCE system will make sure that only the necessary for view interpolation texture and depth are transmitted. When they are received and decoded, the iGLANCE system will ensure view interpolation in order to generate a view from the viewpoint chosen by the user. The new view will contain both texture and depth so that the autostereoscopic screen can generate the other textures necessary for creating a stereoscopic effect.

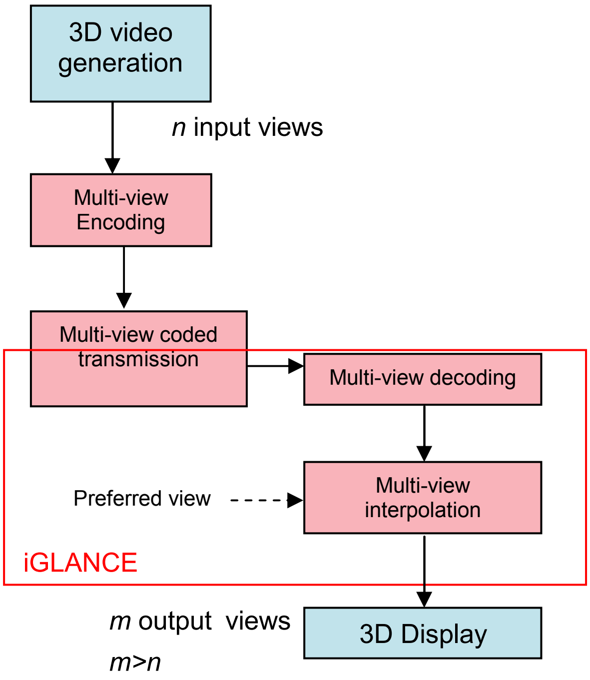


Figure 6. Multi-view transmission chain with iGLANCE decoding and processing.

5. MEDICAL USE CASE

In the case of medical applications, image artifacts due to occlusions may lead to misdiagnosis. Therefore, artifacts should be avoided as much as possible. In order to achieve this, all the multi-view textures are explicitly generated and fed to the stereoscopic display, rather than relying on a single combination of a texture and depth image. Medical 3D imaging will profit from the iGLANCE system by saving bandwidth and time: fewer views will be transmitted compared to the number of views needed for the autostereoscopic display. For each transmitted 3D view (combination of texture and depth), textures are generated only for the missing (multiple) views required by the stereoscopic display. This leads to a lower artifact level, since only a few views are interpolated, so that the occluded areas are inherently much smaller.

Given a number of n views that are being transmitted, and a number of m views that are required by the stereoscopic display, there is multitude of configurations possible, as is illustrated by Figure 7. Since the virtual camera setup can be controlled by the system, the asymmetric right configuration is preferred over the symmetric left one in this example, since it contains fewer interpolated views and will therefore lead to an overall lower artifact level.

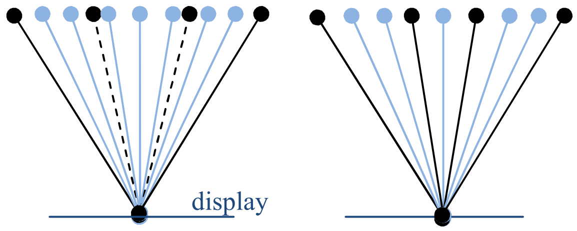


Figure 7. Two configurations for 4 transmitted views, and 9 displayed views. Solid black: transmitted views that can be mapped directly on an output view. Dashed: transmitted views that cannot be mapped on an output view. Light blue: interpolated view.

Lossy compression is an additional source for image artifacts. Both lossy compression and view interpolation allow to reduce the required bandwidth, and the overall system therefore has to be optimized with respect to performance to maintain a low artifact level, combined with moderate bandwidth usage.

Another challenge concerns the latency, which has to be sufficiently low to enable interactive usage in the operating room. At this moment, we define 250 ms latency as the absolute maximum allowed, but the preferred latency is 100 ms.

Contrary to the consumer electronics scenario, the requirements regarding the frame rate for the medical use case are relatively modest; the target frame rate amounts 24 frames per second.

6. CONCLUSIONS

This paper has discussed the use of multiple views of the same scene and the generation of arbitrary in-between views, involving advanced 3D coding and rendering.

The iGLANCE project concentrates on the decoding and rendering of multi-view video with a free-viewpoint and considers also the hardware design of such a receiving platform. Analysis of 3D rendering with multi-view interpolation in a medical use case has shown that it potentially offers lower artifacts and bandwidth reduction.

Working in close collaboration with internationally renowned industrial partners and research institutions, while developing software and hardware simultaneously, the iGLANCE project intends to achieve not only scientific but also relevant results for the industry.

7. REFERENCES

[Ber99] van Berkel, C. Image Preparation for 3D-LCD. Proc. SPIE, Stereoscopic Displays and Virtual Reality Systems VI, Vol. 3639, pp.84-91, 1999.

[Dod97] Dodgson, N.A. Autostereo displays: 3D without glasses. EID: Electronic Information Displays, 1997.

[Mau05] Maupu, D., van Horn, M.H., Weeks, S., Bullit, E. 3D Stereo Interactive Medical Visualization. IEEE Computer Graphics and Applications, Vol. 25, No. 5, pp.67-71, 2005.

[Mor08] Morvan, Y., Farin, D., de With, P.H.N. System architecture for free-viewpoint video and 3D-TV. IEEE Transactions on Consumer Electronics, Vol. 54, pp.925-932, 2008.

[Onu06] Onural L., Ozaktas H.M., Stoykova E., Gotchev A., Watson J. An Overview of the Holographic Display Related Tasks within the European 3DTV Project, Proc. SPIE, Vol 6187, Photon Management II, Strasbourg, France, pp.61870T1-61870T10, 2006.

[Red02] Redert, A., op de Beeck, M., Fehn, C., IJsselsteijn, W., Pollefeys, M., van Gool, L., Ofek, E., Sexton, I., Surman, P. Advanced three-dimensional television system technologies. Proceedings of First International Symposium on 3D Data Processing Visualization and Transmission, pp.313-319, 2002.

[Rui06] Ruijters, D. GPU-Accelerated Volume Rendering to 3D Lenticular Displays. IEEE Workshop on Content Generation and Coding for 3D-Television, Eindhoven, the Netherlands, Vol. 1, 2006.

[Vaz08] Vázquez, C., Tam, W.J. 3D-TV: Coding of Disocclusions for 2D+Depth Representation of MultiView Images. In Proceedings of Tenth IASTED International Conference on Computer Graphics and Imaging (CGIM), Innsbruck, Austria, pp.26-32, 2008.