H-C3 INTEGRATED GRADUATE PROGRAM

Free-Viewpoint Video 3D Camera System

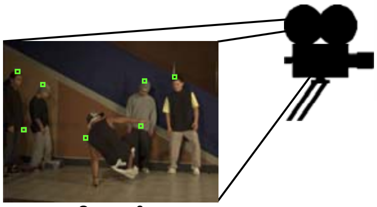
What is free-viewpoint video?

Watching free-viewpoint video will be just like looking out of a window.

Real-time head tracking of the user will let the system know how to render the previously captured scene, thus allowing us to move our head and actually look *behind* objects on the screen. The capability to do so is an important property of human visual 3D perception.

Stereoscopic displays allow the individual generation of two slightly different views – one for each eye. This, together with the free-viewpoint aspect, makes the scene appear almost real, similar to a hologram but dynamic in time.

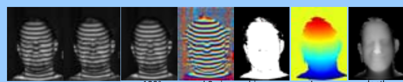
Real world scene [Source: Microsoft Research]



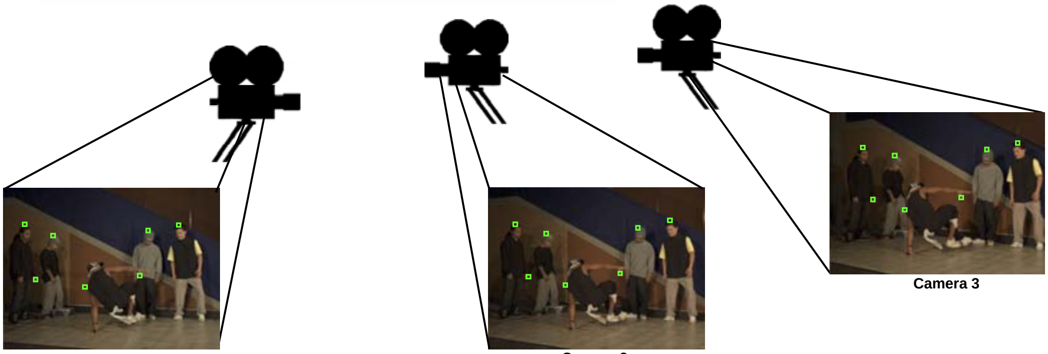
Camera 0

Structured light

The primary optical system consists of two parts: active scan and passive scan. The latter is made up of a minimum of four 2K resolution color cameras. The active scan part consists of two high speed infrared (IR) capable cameras and an IR structured light projector.



The structured light phase shifting technique projects a known sinusoidal pattern onto the scene. In doing so one can compute the so called modulo 2 pi phase map. Noise in this map is detected and removed via a binary mask. A two dimensional phase unwrapping algorithm yields the actual phase, running from 0 to N times 2 pi. As the geometry of the scene distorts the otherwise continuous phase, the gradient of the phase map holds inside the range- or depth map.



Camera 1

Camera 2

Camera 3

Overview

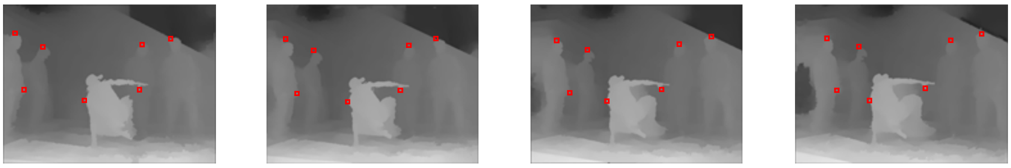
The active scan part of the system is capable of creating high precision depth maps for each video frame. That is, pixel information contains not only color but also distance to camera.

This information is used to generate a 3D mesh, one for each camera. These meshes then undergo noise removal and an alignment procedure, which computes the final global 3D model.

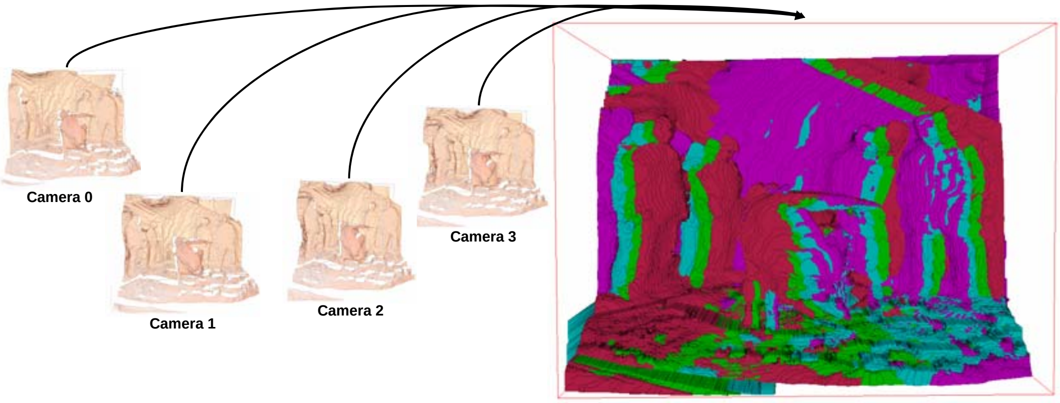
This alignment step computes point correspondences (marked green for color images and red for range images) between the individual camera positions that are in agreement with the epipolar geometry.

Successfully registering the individual 3D meshes into a spatially consistent, global model is crucial for the quality of a single video *frame*. The quality on the actually video *sequence* furthermore strongly depends on temporal consistency between the adjacent 3D meshes that now form the free viewpoint video „frames“.

Additionally, as pixels from the individual cameras do not perfectly overlap in 3D space the free-viewpoint video output resolution is de facto increased by a factor corresponding to the number of cameras in use, say six. This makes the system compatible with the Japanese Ultra High Definition standard, that is a maximum resolution of 7680 × 4320 pixel at an aspect ratio of (16:9).



Camera 0 **Camera 1** **Camera 2** **Camera 3**



About the author

Kai Ide

E-mail: ide@nue.tu-berlin.de

Phone:+49 (0)30 314 - 28 218

Kai Ide is a Ph.D. student within the communications group of Professor Thomas Sikora at Technische Universität Berlin.

He received both his B.Sc. and M.Sc. from Technische Universität Berlin in 2007 and 2008, respectively.

Kai Ide has spent a year in New York, USA as a Rotary International exchange student, and a year in South Korea, where he studied under Professor Kwon at the Quantum Photonics Integrated Circuit Design Lab at POSTECH. He has professional experience in the fields of film- and television, working in film production and as a production assistant for the ARD – German Television at the 29th Olympic Games in Beijing, China.

His current research work is supervised by Professor Thomas Sikora (Communications), Professor Marc Alexa (Computer Graphics), and Dr. Claudio Laloni (Siemens).

His research topic includes the development of a free-viewpoint 3D camera system based on infrared structured light, a system that aims at superseding current stereoscopic systems, used for instance in IMAX 3D movies.

Mesh registration and open questions

As the images above illustrate, the system tries to register a number of 3D meshes, created from different camera viewpoints, into a single, global 3D scene model. This is a cumbersome process and not always successful due to noise, reflections, and other inconsistencies of the meshes.

However, utilizing the precomputed point correspondences one can present a hopefully good enough initial solution to a subsequent Iterative Closest Point (ICP) algorithm that can automatically align given meshes guided by a mean square cost function.

The creation of a spatio-temporally consistent video mesh forms a part of the ongoing research presented here. After holes in the mesh have been completed and after every triangle has been texture mapped other interesting questions arise. How will a human perceive *dynamic* free viewpoint video as opposed to a simple stereoscopic *still* image as can be seen to the right? Will it actually appear real? What possibilities lie in having video together with its 3D model? Can depth image based keying be performed in real-time? These and other unanswered questions make this a very interesting research topic, that has high potential both for academia and industry.