Scientific Report

First name / Family name	Jiang Wang
Nationality	China
Name of the *Host Organisation*	NTNU
First Name / family name of the *Scientific Coordinator*	Leif Arne Rønningen
Period of the fellowship	1/1/2012 to 12/31/2012

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

Collaboration surface is a key function module in the DMP (Distributed Multimedia Plays). The features are stereoscopic, multi-view, near natural size and low latency. The existing infrastructure and off-the-shelf products are FPGA kit, PCIe, microvision Pico laser projector, camera arrays, luminance camera and RGB camera.

From the perspective of collaboration, there are five main parts, e.g. capture stage, encoding, networks transmission, decoding and display stage. My main research is on the 3D capture and display stages. After detail survey of the infrastructure, DIBR (Depth Image Based Rendering) is adopted as the rendering algorithm to generate stereoscopic and multi-view.

The capture stage is important since it provides the source data for processing. In the project, the sparse aperture camera arrays are used. They are strongly or weekly calibrated. To enhance the performance of 3D vision, the texture images and depth images are adopted to generate stereoscopic and multi-view. To get believable rendering effect, the match pixels are statistically analysed. The primitive data structure is proposed during this research. This data is efficient to generate multi-view. Also, the front objects identification and segmentation is important for the networks transmission, since the front objects need high priority in transmission. This strategy is friendly for the protocols developed in our Lab. The protocols treat the front objects with high priority. In the collaboration surfaces project, the collaboration space background could be treated as priori-knowledge. After 3D generation, the qualitative and quantitative performance are tested and evaluated in TESTLIST cases.

To meet the requirement of low latency, the pipeline and parallel design is used in the hardware emulation. The raster scan and two stage memory access strategy guaranteed the high throughput and bandwidth of the abundant vision data. Also, this architecture could guarantee the multi-view buffer fresh rate. Since our expectation of research goal is FPGA or VLSI based system, the architecture design is a bridge between algorithm and implementation. A robust architecture could provide golden data for verification of implementation. And every module partition is key in architecture design, the clean and flexible interface make it easy to be updated.

To meet the near natural size requirements, the microvision Pico laser projector is deployed and studied. This projector MEMS (Micro Electronic Mechanical System) micro mirror is programmable, and the high resolution is available. Not limited to the projector, optic instruments are also used to get 3D multi-view vision. To get autostereoscopic 3D effect, the convex and concave arrays are prepared. The display stage research is still ongoing beyond the end of the ERCIM fellowship.

The top to bottom design methodology applied during the whole research.

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

Real Time Believable Stereo and Virtual View Synthesis Engine for Autostereoscopic Display, 3D stereo MEDIA 2012. Dec. 3-4, 2012, Liege, Belgium

Jiang Wang, Leif Arne Rønningen

Abstract—In this paper, an FPGA and VLSI oriented stereo and virtual view synthesis engine is designed for the autostereoscopic display in the distributed multimedia plays application. To acquire the believable rendering not only for the stereo but also for the distant virtual view synthesis, the DIBR and epipolar geometry are applied in the referenced depth and texture images. The homogeneous depth continuity is statistically analyzed and the corresponding texture is compensated for the hole filling. At the same time, the interpolation and extrapolation are also integrated for the holes which could not be filled where the disoccluded area is not available even in the referenced texture image, either. To meet the real time constraints, computation and memory access intensive modules are realized by hardware accelerating, and the processing is in raster scan pixel line stream rather than frame buffer bulk or macro block line. The simulation and implementation show that based on the correlation of depth in referenced images, the believable rendering is acquired in real time. For the 1920x1080 resolution, the maximum frame rate is 96 fps. This engine is an independent IP and could be integrated with other modules in 3D applications.

III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

ERCIM seminar. Oct. 24-25, 2012, Sophia Antipolis, France

3D stereo MEDIA 2012. Dec. 3-4, 2012, Liege, Belgium

IV – RESEARCH EXCHANGE PROGRAMME (REP)

Sep. 26 – Oct. 4, 2012. Joao Paulo Costeira, <jpc@isr.ist.utl.pt>, Institute for Systems and Robotics, Lisbon, Portugal.

Joao introduced the research team to me. I learned the machine learning, video integral encoding and 3D computer vision theory and technology from this team. Also, I introduced what I did, and got a lot of suggestions and comments from them.

Oct. 15 – Oct. 19, 2012. Roberto Scopigno, <roberto.scopigno@isti.cnr.it>, CNR, Pisa, Italy.

Roberto introduced the research team to me and arranged several researchers to show me what they were engaged in. I learned the high dynamic images and 3D reconstruction theory and implementation from the research team. Also, I had a power point demo to them to show what I did in the computer vision area, and got a lot of suggestions from them.