Scientific Report

First name / Family name	Zahid Riaz
Nationality	Pakistani
Name of the *Host Organisation*	Fraunhofer Institute
First Name / family name of the *Scientific Coordinator*	Rainer Worst
Period of the fellowship	18/07/2012 to 17/07/2013

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

Natural human-robot cooperation in dynamic environments (NIFTi) is an Integrated Project funded by the EU FP7 ICT Programme, in the Cognitive Systems & Robotics unit (Project #247870). NIFTi runs from January 1 2010 until December 31 2013.

This report refers to the research work conducted at Fraunhofer Institute in the context of NIFTi project during the fellowship program. This project investigates how a cognitive robot can complement models of its own capabilities and situation awareness, with cognitive user models of task load and workflow. It works off the hypothesis that a combination of these can achieve natural human-robot cooperation. The idea is to continuously balance how to operate autonomously, with how to cooperate so as to minimize task load for the human, and optimize workflow.

Besides different experiments in the domain of environmental mapping using ICP algorithms, simultaneous localization and mapping (SLAM) algorithms and cooperative mapping between multiple robots, one of the concrete output of this fellowship was an optimal solution for such scenarios using 3D reconstruction from single view.

For the autonomous navigation of the robots in unknown environments, generation of environmental maps and 3D scene reconstruction play a significant role. SLAM algorithm helps the robots to perceive, plan and navigate autonomously whereas scene reconstruction helps the human supervisors to understand the scene and act accordingly during joint activities with the robots. For successful completion of these joint activities, a detailed understanding of the environment is required for human and robots to interact with each other. Generally, the robots are equipped with multiple sensors and acquire a large amount of data which is challenging to handle. In this regard an efficient 3D scene reconstruction approach is propose for such scenarios using vision and graphics based techniques. This approach can be applied to indoor, outdoor, small and large scale environments. The ultimate goal of is to apply this system to joint rescue operations executed by human and robot teams by reducing a large amount of point cloud data to a smaller amount without compromising on the visual quality of the scene. From thorough experimentation, we showed that the propose system is memory and time efficient and capable to run on the processing unit mounted on the autonomous vehicle. For experimentation purposes, we use standard RGB-D benchmark dataset.

For a robot-centric rescue operation, a comprehensive situational awareness is a crucial step for the success of robotic aided search and rescue missions. The human operators need to make decisions which are based on the information directly derived from the situation. The research has shown that the quality and reliability of these decisions depend on the impression of situational awareness rather than the amount of data provided to the human, i.e. stakeholders do not need all available data but enough data to assess a situation.

Moreover in rescue scenarios, some of the areas are not accessible to humans and are suitable for the robots to access and convey useful information to the person remotely connected with them. In such situations robots can be helpful for the rescue team members if they can access these areas (like tunnels, fire-caught areas) and can efficiently transmit every instance of the sensory data in realtime. In USAR scenarios, a combination of unmanned ground vehicle (UGV) and unmanned aerial vehicle (UAV) acquires a large set of laser scan data and videos of the operational area. Since these robots are connected with each other and to their human team members through a limited transmission bandwidth which may degrade during a field operation. The robot needs to adjust to the communication capacity at each point in time. In order to make this communication possible between human-robot team member, it is however required to reduce the amount of the data without compromising on the quality of the useful information. For such situations, we propose a memory and time efficient approach to process this data on the onboard processing unit of the vehicle. The goal is to carefully calculate the amount of data which is sufficient to perceive such scenarios and adequate to transmit over the limited bandwidth in real time. We benefit from efficient algorithms which can run in real time on the PC mounted on the robot and reconstruction of the scene can remotely be performed on ordinary CPU instead of using high quality GPU. The final visual quality, memory and time efficiency are not compromised by introducing sparseness in the point cloud data. The approached followed in this research work is shown in Figure 1 whereas Figure 2 shows the result of the scene reconstruction from a single view.

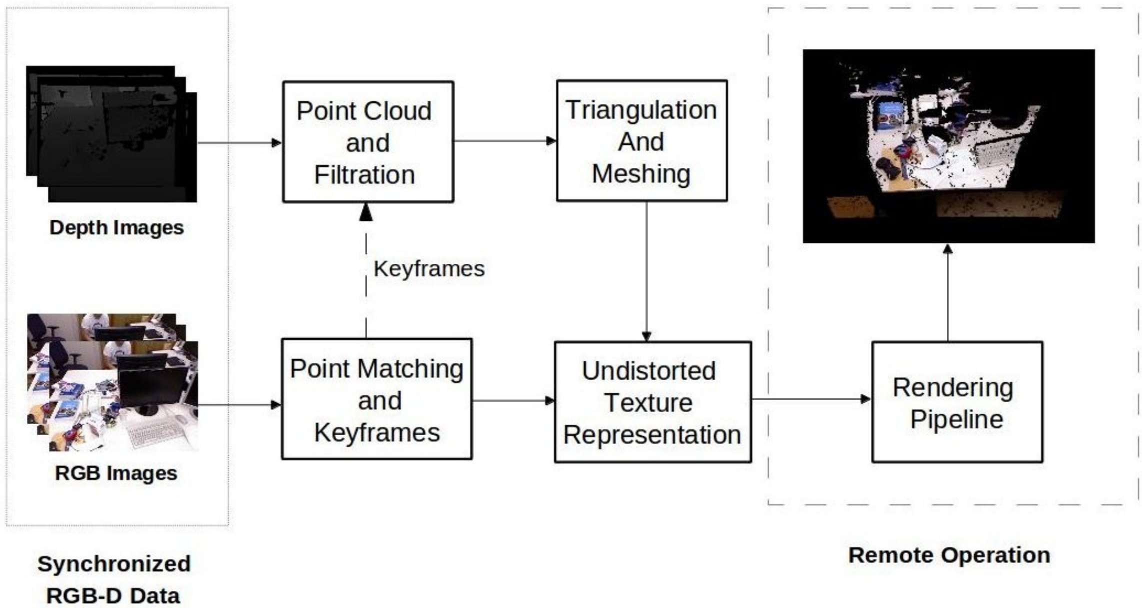

Figure 1: Overview of our approach. The input to the system is depth image sequence (or point clouds) and synchronized RGB images. The output is 3D reconstruction of the environment which is efficient in computing, storage and capable to apply in real time

The solution is devised by using an intermediate vision and graphics based approach to generate real time 3D model of the environment. Generally, a laser scanner mounted on the robot acquires scans at 30Hz and needs to transmit it to the remote station. For example, a Kinect sensor acquires a point cloud of size more than 300K points at 30Hz. This data is reduced to a few thousands of points describing a large area visible to the robot. Since this reduced point cloud contains a small number of points which are not sufficient to visualize the scene by rendering each point. We render its triangular surface mesh. For a few thousands of points 3D reconstruction can be performed on remote computer by acquiring downsampled point clouds and corresponding RGB images directly transmitted by the robot. A dense point cloud with RGB values provides a detailed representation of the scene but amount of data is large and requires more memory to store and more time to transmit and process this data. On the other hand, with a sparse point cloud representation to a few thousands meaningful points, it is challenging to attain a high visual quality. We achieve a better texture quality by generating an undistorted texture map of the environment and efficiently rendering the triangular mesh. In this way, we achieve a better final visual representation of the scene. The results obtained on standard dataset are given in Table 1.

Figure 2: Result of 3D scene reconstruction from a single view by using downsampled data from RGBD sensor.

In this regard, we have proposed an efficient approach to reconstruct an environment while simultaneously reducing the amount of data while preserving the visual quality of the scene. The purpose of this approach is to apply it in robotic aided USAR scenarios where human and robot teams are performing different activities together. Since the amount of data acquired by the robots is large, we focus our attention to reduce the size of the data while not compromising on the quality of data. In this way, we provide a solution for situational awareness by reducing the amount of RGB-D data required to transmit from robot to user. We evaluated and compared the approach using a standard RGB-D benchmark dataset and achieved a downsampled representation up to (5-12)% of the original cloud while keeping the visual quality still understandable for human users. For future work of this project is to make usage of this approach for real world robotic aided search and rescue operations.

	No. of points	No. of triangles	Down-sample time (sec)	Size of final map (KB)	Rendering time (sec)
Original Cloud	307200	406533	NA	9395.5	0.27
Our Approach	22474	40168	0.03	8.5	0.03

Table 1. Average values over eight random point clouds from freiburg1 dataset. The figures shows different values of the key steps used in this paper. A comparison to full point cloud to reduced point cloud is given.

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

1. Zahid Riaz, Thorsten Linder, Sven Behnke, Rainer Worst, Hartmust Surmann, “Efficient Transmission and Rendering of RGB-D Views”, International Symposium of Visual Computing (ISVC), Rethymnon, Crete, Greece, 29-31 July 2013.

III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

1. NIFTi joint exercise (NJEx), 3-5 October 2012, Dormund Germany
2. NIFTi-Vollversammlung, 20-21 March 2013, Saarbrücken, Germany
3. International Symposium of Visual Computing (ISVC), Rethymnon, Crete, Greece, 29-31 July 2013.

IV – RESEARCH EXCHANGE PROGRAMME (REP)

Nil