

# 3D Building Reconstruction and Thermal Mapping in Fire Brigade Operations

Christian Schönauer  
schoenauer@ims.tuwien.ac.at

Emanuel Vonach  
vonach@ims.tuwien.ac.at

Georg Gerstweiler  
gerstweiler@ims.tuwien.ac.at

Hannes Kaufmann  
kaufmann@ims.tuwien.ac.at

Vienna University of Technology

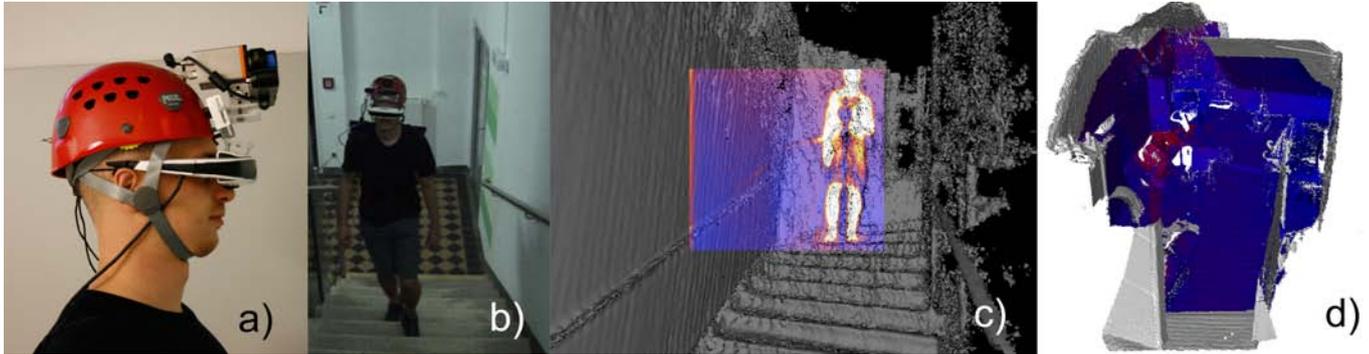


Figure 1: a) Our setup featuring a depth sensor, thermal camera and head mounted display b) connected to a notebook for mobile operation. c) Real-time reconstruction of the staircase with an overlay of the thermal image. d) Mesh reconstructed from the structure of a room textured with the images of a thermal camera.

## ABSTRACT

Fire fighting remains a dangerous profession despite many recent technological and organizational measures. Sensors and technical systems can augment the performance of fire fighters to increase safety and efficiency during operation. An important aspect in that context is the awareness of location, structure and thermal properties of the environment.

This paper focuses on the design and development of a mobile system, which can reconstruct a 3d model of a building's interior structure in real-time and fuses the visualization with the image of a thermal camera. In addition the position and viewing direction of the fire fighter within the model is determined and a thermal map can be generated from the gathered data, which could help an operational commander to guide his men during a mission.

First tests with our system in different situations showed good results, being able to reconstruct different larger scenes and create thermal maps thereof.

## Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities

## Keywords

Real-Time Dense Reconstruction, Thermal Camera, Augmented Reality, Fire Fighter Safety

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AH'13, March 7–8, 2013, Stuttgart, Germany.

Copyright 2013 ACM 978-1-4503-1904-1/13/03 ...\$15.00.

## 1. INTRODUCTION

The Center of Fire Statistics reports that over hundred fire fighters lost their lives in the line of duty in addition to close to four thousand fire deaths per year in the US alone (data from 2003 and 2004 [3]). These numbers suggest much room for improvement through technical systems in that field. The safety and efficiency of firefighters during operation depends to a large degree on knowledge of location within a building and the thermal situation. Only then it is possible to guide them towards targets or out of critical situations. Therefore, tracking the position is an important objective of our system intended to support fire fighters during a mission. In addition to the position in space, it is important for an operational commander to know about the structural context (i.e. building structure) in order to safely direct his men. Accurate maps of buildings are rarely available before deployment. Therefore, our system follows an approach reconstructing the 3d structure of a building in real-time [7]. We have combined the 3d reconstruction with thermal imaging in order to visualize thermal properties for fire fighters (Figure 1c) and generate thermal maps (Figure 1d). These maps could then be used to provide an operational commander with a better overview of the situation highlighting hot spots or human victims in need of rescue.

Therefore, our **contributions** include:

- Development of a **mobile setup** (Figure 1a/b) for 3d reconstruction and thermal imaging.
- **Fusion of a 3d reconstruction algorithm with thermal imaging** for:
  - Real-time visualization of building structure and thermal overlay.
  - Reconstruction of a 3d thermal map color coding interesting temperature ranges.

## 2. RELATED WORK

State of the art approaches give a solution to the problem of self-localisation and reconstruction in unknown environments by realizing Simultaneous Localisation and Mapping (SLAM) with visual data and additional sensors. SLAM has undergone great advances in the last decade [2, 10].

Depth cameras provide input to a new class of mapping and tracking algorithms e.g., [7]. Here, the depth information is not calculated from 2d images, but directly streamed from a sensor. Existing depth cameras deliver a 2.5d picture and are based on one of three different technologies: (1) Time-of-Flight: Based on the measurement of the time it takes a modulated light signal to be projected, reflected in the scene and returned to the sensor. (2) Structured Light: A structured light pattern is projected and its distortion in a camera image is used to triangulate depth. This is the operation principle of the sensors developed by Primesense, e.g. Asus Xtion PRO [1]. (3) Light-field Camera: A plenoptic camera can be used to calculate depth information.

Kinect Fusion is an algorithm for reconstructing 3d environments and tracking a camera's pose at the same time [7]. The algorithm requires a continuous stream from a depth camera. It allows for real-time dense volumetric reconstruction and was originally developed for smaller workspaces providing reconstructions with an accuracy down to a few millimeters at close range. In a recently published technical report [12] the mapping and tracking based on Kinect depth data for larger indoor environments has been introduced. In parallel, an extension to the Kinect Fusion algorithm (KinFu Large Scale) has been implemented in the Point Cloud Library (PCL) for the same purpose [5].

Cameras that generate thermal images in the long wave infrared range (LWIR) are referred to as thermal cameras or FLIR (forward-looking infrared) cameras. Each object (surface) with a temperature higher than absolute zero ( $0^{\circ}\text{K}$ ) emits energy with a certain wavelength. Cameras operating in the LWIR band (i.e. radiation with wavelengths in the range of 8-15  $\mu\text{m}$ ) capture most of the thermal radiation that we are confronted with in our natural environment. In our system we employ uncooled cameras because they are cheaper, smaller and consume less energy than cooled cameras, which makes them well-suited for head or helmet-mounting. A description of head-mounted thermal cameras in the context of firefighting is provided by [6]. The temperature measurements of uncooled cameras are influenced by factors, like sensor temperature, which need to be compensated for. Thermal cameras are usually radiometrically calibrated to allow for proper temperature assessments. For a calibrated camera each pixel's gray-value corresponds to an absolute temperature. This is important for many tasks in the context of the firefighting domain, e.g. detection of body temperature for finding people. However, an uncooled camera produces reliable measurements only if calibrated to a small temperature range (e.g. from 0 to  $100^{\circ}\text{C}$ ). This limits usefulness for a fire fighting scenario, with high temperature dynamics. Consequently, we used the camera also without calibration. Uncalibrated cameras give only relative temperature assessments. The gray-value of a pixel cannot be mapped to a particular temperature (or amount of emission). However, due to the enhanced contrast an uncalibrated camera can serve better in visualizing building structures, which can be helpful in scenarios where a fire fighter faces poor visibility in the visual spectrum.

## 3. SETUP AND HARDWARE

In our setup (Figure 1 a/b) we use an Asus Xtion PRO [1], which provides depth data using a the structured light approach described in section 2. The depth data is streamed at a resolution of 640x480 pixel and a frame rate of 30 frames per second. For thermal imaging we employ the Xenics Gobi-640-GigE camera. It provides a resolution of 640x480 on an uncooled microbolometer array with a  $17\mu\text{m}$  pixel pitch at 50 Hz and a peak sensitivity at a wavelength between 8 and 14  $\mu\text{m}$ . The camera provides a thermal resolution of  $0.05^{\circ}\text{C}$  and is specified for an ambient temperature of  $-40^{\circ}\text{C}$  to  $70^{\circ}\text{C}$ . For our test we had a lens available with a focal length of 18 mm and  $30^{\circ}$  horizontal and  $23^{\circ}$  vertical field of view (HFOV/VFOV), which is significantly less than the Asus Xtion ( $57^{\circ}$  HFOV and  $43^{\circ}$  VFOV). Therefore, the sub-image of the depth reconstruction covered by the thermal camera is relatively small, as illustrated in Figure 1c. The thermal camera has dimensions of 49 x 49 x 70  $\text{mm}^3$  and a weight of 220 g without lens. For our tests we used both, the standard calibration for a temperature range from  $-20^{\circ}\text{C}$  to  $100^{\circ}\text{C}$  and the automatic calibration with relative temperature values. Our mobile setup (Figure 1b) was based on a notebook with an Intel Core i7-3720, 16 GB main memory and the NVIDIA GeForce GTX 680M with 4096 MB RAM. For the visualization we used a Silicon Micro Display ST1080 Head Mounted Display.

## 4. SYSTEM OVERVIEW

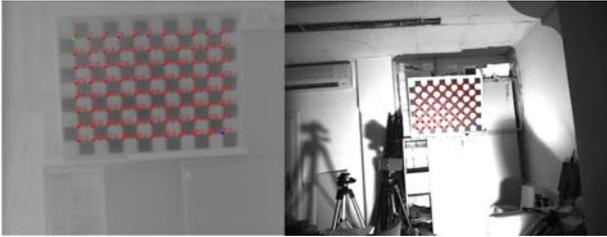
For 3d reconstruction we have used the Kinect Fusion algorithm [7] and its extension for large scale environments as implemented in the Point Cloud Library (PCL) [5]. We have integrated the thermal camera in the PCL framework and extended it to create thermal visualization and a thermal map instead of texturing with RGB images as in [5].

### 4.1 Calibration and registration of the thermal camera with the depth sensor

In order to present an overlay of a thermal image over the 3d reconstruction and creating a thermal map, we needed a precise alignment or registration between the thermal camera and the depth sensor. Therefore, we calibrated the cameras' intrinsic parameters as well as their extrinsic transformation relative to each other. For the calibration of the depth sensor the laser projector was covered and only the images of the near-infrared camera were used. The thermal camera was utilized without temperature calibration, since only the image geometry is important in this context. We took approximately 20 images in the expected main working range of 2.5 to 4 meters simultaneously with both cameras using a self-developed tool triggering them at the same time. The thermal images are automatically inverted in our tool, so the intensities match with the depth sensor's camera. In the next step we utilized the MIP Multi Camera Calibration tool [11] to calibrate intrinsic and extrinsic parameters of both cameras. We used an A0 sized checkerboard pattern illuminated by a 1000 Watt halogen spot. The black squares of the pattern heated up stronger resulting in a sufficiently distinguishable thermal pattern (See Figure 2).

### 4.2 Kinect Fusion Algorithm

Kinect Fusion processes every frame obtained from the depth camera and generates three things: an updated 3d



**Figure 2: Images taken for calibration from the thermal camera with intensity inverted (left) and from the Asus Xtion's near infrared camera (right). Note the bright illumination and the difference in the pattern's sizes originating from different FOVs.**

structure containing information on surfaces in the scene, the camera pose and a rendered image from the perspective of the camera. First from the 2.5d depth image the 3d positions and normals are calculated in a preprocessing step. Using the Iterative Closest Point (ICP) method the pose of the current frame is aligned with previous data contained in a 3d structure - the volumetric Truncated Signed Distance Function (TSDF) [4]. Once aligned, information from the current depth frame is merged in the TSDF volume. Data from multiple frames is averaged in the depth/surface values of the TSDF structure, which is therefore more accurate than a single depth image and contains less noise and holes. In addition, an image can be rendered in real-time by ray-casting the TSDF. Finally, in a postprocessing step a mesh can be extracted from the TSDF volume (Figure 3a).

The TSDF volume is kept in GPU memory for fast and parallel processing [7]. This, however, limits the size of the scene that can be reconstructed to the volume of a small room. To avoid these limitations areas of the TSDF structure, which are not currently used for pose estimation, are moved to main memory as implemented by [5]. This in turn allows new parts of the scene to be reconstructed in GPU memory, allowing a more extensive work space.

### 4.3 Implementation

Our implementation of the Kinect Fusion algorithm has been developed in C++ and is based on the open-source implementation of the Point Cloud Library (PCL) [5, 9] in large parts. We have made use of the available methods implemented in CUDA, optimized for processing on the GPU. We access the thermal camera via the available SDK and preprocess the images making extensive use of OpenCV [8] image processing library. In addition, we have implemented a customized Grabber module to stream images into the PCL framework. The frame rate of the 3d reconstruction (15Hz) is lower than the thermal imaging (50Hz). Therefore, we discard thermal images that do not precisely match the depth frames for further processing. For the live view the image from the ray-traced TSDF is overlaid with the current thermal camera picture. We have chosen the rendering of the TSDF as underlying display mode, because it works in low light conditions (as opposed to the RGB camera) and offers higher accuracy than a single depth image.

The application-module for reconstruction is configured for a cubic TSDF volume with six meters side length and separated in 512 steps. This is sufficient for most of our purposes and we can cover staircases or other larger vertical structures. A larger volume would only make sense in

combination with a depth sensor also performing at a larger range as discussed in section 6.

In our current prototype meshes are generated in an offline process using the marching cubes algorithm [5]. The meshes are then textured using thermal images taken during the reconstruction process. For our test sequences we kept two thermal images per second for texturing. When images are stored for later processing, the poses of the depth sensor are also logged. Using the calibration between the depth sensor and the thermal camera as well as the intrinsic parameters of the thermal camera, we project the thermal images onto the mesh and generate textured coordinates from the projections as described in [5]. The result can be seen in Figure 3e. While in the current implementation these maps are available only after the reconstruction we plan to make this step more interactive, as described in section 6.

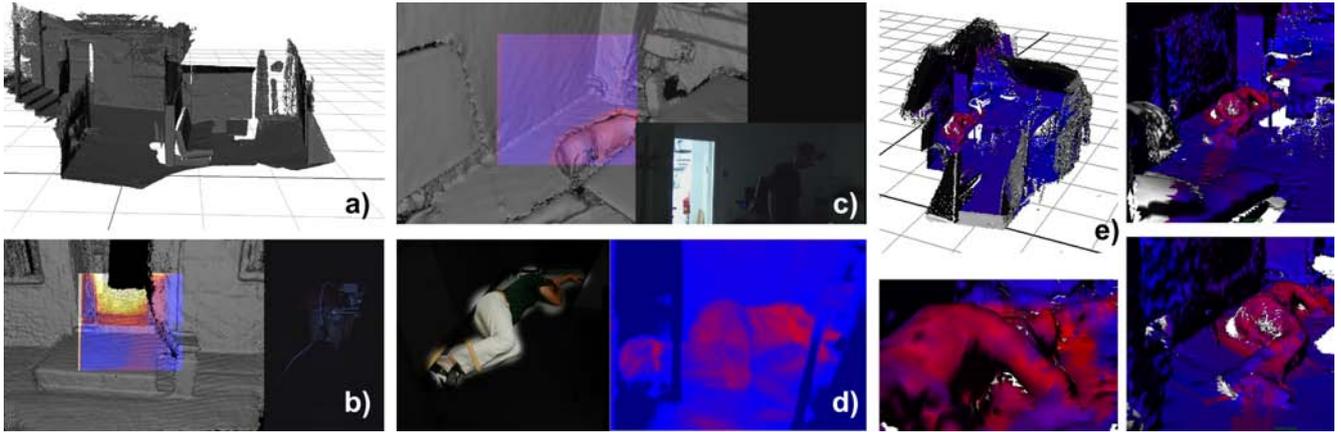
## 5. EVALUATION

First tests with our system in different situations showed good results. We were able to reconstruct different large scenes like a building entrance, different rooms, hallways and staircases (Figure 3) with a level of detail, which should be more than sufficient for the proposed purpose. Reconstruction also works in low light conditions or complete darkness and can therefore help a user navigating in such conditions as illustrated in Figure 3b. Thermal imaging with calibrated temperatures enabled us to clearly highlight human body temperature in red in environments with room temperature (See Figure 3c-e). Our experiments with thermal imaging depicting relative temperatures proved useful in situations with smaller temperature differences (i.e. fine structures) or outside the calibrated temperature range (See Figure 3b).

Reconstruction with thermal imaging showed good registration of the cameras. Only situations where tracking lacks accuracy sometimes cause misalignment. This is due to the fact that the tracked pose is used to project a thermal image onto the reconstructed mesh. Our mobile setup performs reconstruction and tracking at an interactive rate of 12-15 frames per second. Thermal imaging does not induce a noticeable overhead in the processing loop.

## 6. FUTURE WORK AND LIMITATIONS

Reconstruction with our setup could aid a user in a fire fighting operation by mapping building structures up to the proximity of a fire source. However, we have not yet tested our system in dense smoke or environments with high temperature variation. From the technologies used, it has to be expected that performance will be significantly hampered by anything blocking or degrading the quality of a direct view in the visual and especially in the near-infrared spectrum. In environments densely covered with smoke it is very likely that reconstruction is brought to a halt. In this case a user could, however, still revert to LWIR imaging to retrieve information about structures and people in the building. In the future a depth camera also working in the LWIR band could reduce this limitation. For deployment in hostile environments the setup's hardware components have to be sufficiently robust and able to work in high ambient temperature. While thermal cameras are already being used for higher temperature ranges in fire fighting operations, this is still an engineering challenge for other parts.



**Figure 3:** a) Mesh reconstructed from the point cloud generated by the Kinect Fusion algorithm. b) Live reconstruction and thermal image overlay with low light conditions at night and c) in a simulated search and rescue scenario. d) Thermal imaging without reconstruction highlighting human temperature range (in red). e) Mesh textured with thermal images shown from different perspectives.

The Kinect Fusion algorithm by design requires a platform, which is capable of high performance parallel computations. Currently, this is only possible on a fast GPU, hampering mobility. However, development of fast GPUs for mobile devices lets us expect a lighter solution for the near future. Furthermore, our implementation is currently restricted to slow movements. However, optimizations in the implementation and future hardware developments are expected to solve this issue. Asus Xtion and other devices with similar technology provide data useful for our purposes at a range of up to five meters. In future work we plan to integrate a light field camera into our system, reducing this limitation.

Currently, the creation of the mesh and texturing is handled in an offline process taking approximately one minute for an average scene. In future work we plan to send regular updates from the mobile setup to a command station. Then the mapping process can be parallelized with the reconstruction. In addition a commander provided with a close to real-time update of the situation. Merging reconstruction from multiple users and sustaining communication in hostile environments provide interesting research topics.

Another aspect not yet sufficiently explored is user interfaces and information visualization for fire fighters and an operational commander. This offers a rich field of research with (1) different display modes to be assessed for each situation, (2) the available structural and thermal data to be analyzed for automatic detection of fire sources or human victims and (3) interaction with the system during operation.

## 7. CONCLUSION

In this paper, we presented our mobile system for 3d reconstruction and thermal imaging targeting fire fighter operations. It can provide real-time visualization of a buildings interior structure and thermal situation to a fire fighter as well as a three dimensional thermal map for an operational commander. First tests showed good results, while limitations and future directions have been identified.

## References

- [1] Asus. Asus Xtion PRO. URL [http://www.asus.com/Multimedia/Motion\\\_Sensor/Xtion\\\_PRO/](http://www.asus.com/Multimedia/Motion\_Sensor/Xtion\_PRO/). 2013.
- [2] T. Bailey and H. Durrant-Whyte. Simultaneous localization and mapping (SLAM): part II. *IEEE Robotics & Automation Magazine*, 13(3):108–117, Sept. 2006.
- [3] N. Brushlinsky, J. Hall, S. Sokolov, and P. Wagner. World wide fire statistics. Technical report, Center of Fire Statistics, 2008.
- [4] B. Curless and M. Levoy. A Volumetric Method for Building Complex Models from Range Images. *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 303–312, 1996.
- [5] F. Heredia and R. Favier. KinFu LargeScale. URL <http://www.pointclouds.org/blog/srcs/fheredia/index.php>. 2012.
- [6] J. Millerand and H. Duvoisin. Applications and Performance of IR Helmetcams. *Storming Media - Pentagon Reports*, 2009.
- [7] R. A. Newcombe, A. J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. *10th IEEE International Symposium on Mixed and Augmented Reality*, pages 127–136, Oct. 2011.
- [8] OpenCV. Open Source Computer Vision. URL <http://opencv.org/>. 2013.
- [9] PCL. Point Cloud Library. URL <http://pointclouds.org/>. 2013.
- [10] G. Reitmayr, T. Langlotz, D. Wagner, A. Mulloni, G. Schall, D. Schmalstieg, and Q. Pan. Simultaneous Localization and Mapping for Augmented Reality. *2010 International Symposium on Ubiquitous Virtual Reality*, pages 5–8, July 2010.
- [11] I. SCHILLER. MIP - MultiCameraCalibration. URL <http://www.mip.informatik.uni-kiel.de/tiki-index.php?page=Calibration>. 2011.
- [12] T. Whelan, M. Kaess, M. Fallon, H. Johannsson, J. Leonard, J. McDonald, and J. J. Leonard. Kintinuous : Spatially Extended KinectFusion. Technical report, MIT, Boston, MA, USA. URL <http://dspace.mit.edu/handle/1721.1/71756>, 2012.