Beyond 3D Imaging and Visualization

Lena Klasén[*†], Amanda Berg[‡§], Per-Erik Forssén[‡] and Haibo Li[¶]
[*]Swedish National Police Authority, Sweden
[†]Information Coding Group, Linköping University, Sweden
[‡]Computer Vision Laboratory, Linköping University, Sweden
[§]Termisk Systemteknik AB, Linköping, Sweden
[¶]Media Technology, Royal Institute of Technology, Sweden
Email: {lena.klasen, amanda.berg, per-erik.forssen}@liu.se, haiboli@kth.se

***Abstract*—Visual Sweden is a national Swedish initiative that promotes innovation and regional growth within the visualization and computer vision fields. This paper presents a project within Visual Sweden that further develops today's high-resolution, precise 3D models into unique and intelligent n-dimensional models for analysis and visualization of complex environments and events. We present the project scope and give some examples of outcomes based on data from a measurement campaign conducted in June 2019 in Västervik, Sweden. The overall goal for the project is to provide means for new industrial products and services plus tools and methods for governments, whilst feeding back ideas for further research topics to academia. This is achieved by creating a platform for exchange of expertise, technology and algorithms. The project includes a broad set of participants from governments, academia, research institutes and industries, representing various parts in an innovation value-chain, from end-user to market. Together they secure a good knowledge of end-user's and market needs.**

I. Introduction

3D modeling technologies are nowadays commonly used to an increasing extent in various applications. 3D models can be developed from camera data, laser radar data, thermal imaging and ground penetrating radar. One application example is the use of a 3D model in forensics. To understand the course of event at the terrorist attack at Drottninggatan in April 2017, Swedish National Forensic Centre (NFC) digitised the entire street constituting the crime scene together with a truck, illustrated in Fig. 1. The high accuracy 3D model and the possibility to reconstruct the event over time, enables visualisation and analysis of the course of events in the large and complex crime scene. This includes visualizing the event as seen by the perpetrator when driving the truck, Fig. 2.

However, there is often more information available to describe, analyse and visualize a scene that can be incorporated in a digitised model. For example, additional sensor types add more dimensions to a model than spatial information and color. This kind of added functionality can provide new capabilities to handle more complex environments and events and also increased efficiency in daily operations. In this project we expand the use of 3D models into unique and intelligent n-dimensional models (n-D models) for analysis and visualization of complex environments and events.

A key issue is to establish a platform for collaboration that encourages exchange of knowledge, development of algorithms, exchange of technologies and the introduction of new sensors. A core activity is facilitation of method refinement and adaptation to specific needs. Hence, we combine multidimensional sensor informatics and intelligent algorithms into intelligent multidimensional imaging. This concept offers unique access to a laboratory environment that provides both knowledge, sensors and algorithms with the possibility of imaging a scene in more dimensions. Due to the availability, cost and complexity of implementation, a similar environment can be very difficult to create for an individual organization or company. In addition, the project partners holds a good knowledge of end-user and market needs. The goal is to provide the project partners with knowledge to be able to bring the results to the market within a reasonably short time frame.

Fig. 1. 3D model of Drottninggatan, Stockholm, Sweden. The model is from registration of many terrestrial lidar scans.

II. n-dimensional modeling

Sensors that are sensitive to radiation in wavelength bands outside the visible spectrum, provide additional important information. For example, adding thermal imaging to see heat radiation in forest fires or spectral imaging to single out body fluids from other substances in a crime scene. Radar sensors allow underground areas to be mapped or buildings "looked into", by providing the ability to penetrate material such as tarmac, gravel and walls. Multi-sensor set-ups, along with different types of metadata (weather, geography, images from social media, etc.) are used to further develop tools and methods for documentation, analysis, interpretation and visualization of a measured scene and the course of event.

Fig. 2. 3D model of Drottninggatan as seen from the driver, Stockholm, Sweden. The model is from registration of many terrestrial lidar scans.

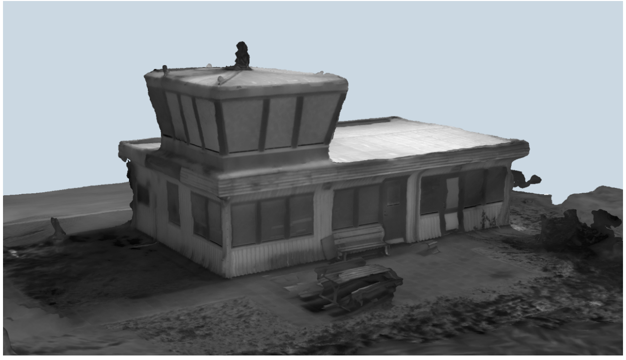

Fig. 3. A 3D model of the airport control tower, created through Spotscale's Structure from Motion Multi-View Stereo pipeline using only images from the FLIR T1030 camera. Brighter colors represent warmer temperatures and darker represent colder.

Depending on the application at hand, intelligent algorithms are required for analysis and visualization of the large amounts of sensor data and metadata, spanning from a need for low resolution real-time to high resolution off-line processing. As the sensor- and metadata provides heterogeneous information, there is work on pre-processing and synchronization preceding further analysis of the data in a multidimensional context. For example, the analysis may consist of intelligent algorithms for detection, tracking or recognition and the results visualized for an end user in scene analysis such as crime scene investigation or finding toxic chemicals, object detection and location, mission planning or decision support.

Continuous-time 3D-model estimation from cameras and lidar sensors is used in the project. Both lidar and consumer cameras (e.g. cellphones) are scanning sensors (in video this is known as rolling shutter). Scanning acquisition leads to geometric distortions in 3D-models if they are acquired during sensor motion, while the registration or model fusion assumes stationary capture. Continuous-time modelling instead correctly models the sensor pose as a trajectory, e.g., using a spline. One benefit of continuous-time modelling is improved accuracy for lidar models acquired from rooftop sensors on moving cars. Another is that it can allow the use of ad-hoc video filmed by passers-by during an accident or a crime. Such video is almost exclusively filmed with mobile phones with rolling shutter. The Computer Vision Lab (CVL) at Linköping University conduct basic research on continuous-time estimation [1] and registration [2].

Another area of special interest is to explore means for finding what part of large amounts of information leads to desired results. This is, for example, an important area in research on autonomous vehicles to figure out what probability there is for a pedestrian to appear in a certain situation to avoid decisions based on speculative data. Furthermore, it is of interest to explore the applicability and confidence in the result when methods of deep learning are applied. The physical environment could support or restrict certain behaviors of humans and objects in the scene. The environment and behaviors have a reciprocal relationship, which is known as *affordance* defined by J. Gibson [3]. Computing spatial affordances from the reconstructed n-D environment is an efficient and effective way to characterize how people develop sensory integration and how people adaptively respond through various sensory-related actions.

III. Västervik Measurement Campaign

For the purpose of collecting data for demonstrating the project's concept, a measurement campaign was conducted June 20 - 21, 2019 in and in the vicinity of Västervik, Sweden. The airport and city centre were selected and a broad set of sensors were used, spanning from satellites to underground and underwater imaging. Several project participants collaborated, for example, by using one partner's drone to carry another partners sensor and a third company to process and visualise the data. Here we only give a few examples of how the Västervik data is used.

Termisk Systemteknik AB recorded sequences using a thermal infrared camera, FLIR T1030, mounted on a drone operated by Mainbase AB. In addition to the sensor, the drone was also equipped with a GPS. Due to flight restrictions in urban areas, data was only collected at the airport. An example of a 3D model built by Spotscale AB from collected thermal infrared images depicting the airport control tower can be seen in Fig. 3.

Researchers from CVL have recorded data from an Ouster OS-1 64 line lidar, a GoPro Hero 6 action camera, as well as an Occam Omni-stereo panorama camera. The lidar sensor records scans at 10Hz, at 2048 rotation angles, see Fig. 4 for an example of output at Grönsakstorget in Västervik. Here, 380 lidar scans have been fused using the *registration loss learning* (RLL) algorithm [4] which was developed in the project.

Glana Sensors AB contributed to the measurement campaign with their own hyperspectral camera. The sensor is built using a COTS CCD image sensor, on which a continuously variable bandpass filter is mounted, enabling the collection of spatio-spectral images. That is, the wavelength of the captured radiation varies gradually over the columns of the image. Wavelengths range from 450 to 950 nm. During the campaign, Glana collected hyperspectral images of the forest and airport control tower from a drone provided by RISE SICS East. A 3D model of the airport

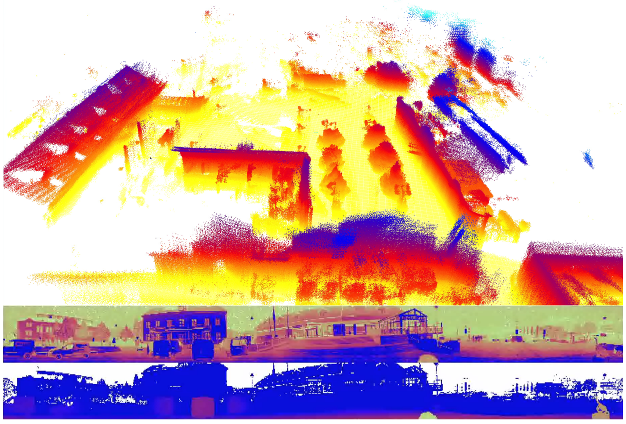

Fig. 4. Model of Grönsakstorget in Västervik from a sequence of hand-held Ouster OS-1 64 line lidar scans. Top: 3D model, computed by unpacking 380 range scans to point clouds and registering them using RLL. Colour from elevation. Middle: Single scan reflectance map (i.e. range compensated intensity, 64×2048), Strong to weak reflectances are coloured in a blue-to-yellow colour map. Bottom: Single range scan (also 64×2048 red is close blue is far away).

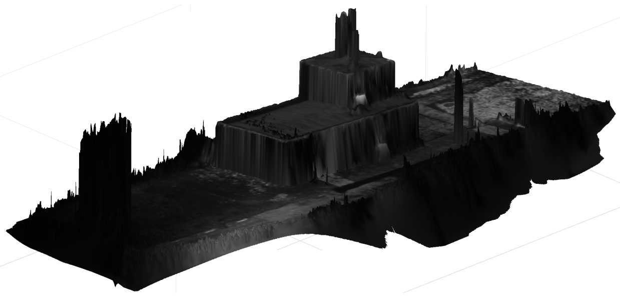

Fig. 5. 3D model of the airport control tower using images from the Glana hyperspectral camera.

control tower built from the hyperspectral images can be seen in Fig. 5, and an example from the forest is provided in Fig. 6.

NFC set up a scenario illustrating a murder case, using car, a dummy and a variety of items representing forensic evidence. The scene was recorded by Spotscale AB using conventional cameras and photogrammetry, seen in Fig. 7, and by NFC using lidar data, seen in Fig. 8. The "crime scene" was also recorded for further analysis by other sensors, for example using MainBase's drone, Termisk Systemteknik's IR-camera and Glana's hyperspectral camera mentioned above.

Media Lab at KTH addresses another challenging question in the field, how to anticipate human intention from the extracted and reconstructed 3D models of the environments. Based on the arguments on the existing approaches an alternative approach to the problem based on Heidegger's and Merleau-Ponty's phenomenology [5] and J. Gibson's ecological psychology [3] is suggested and explored by Media Lab. The starting pointing is from the thinking of how to infer human intention from human body or body gestures. Instead of directly working on the general problem of the estimation of human intention, a concrete, real world problem in the field of autonomous

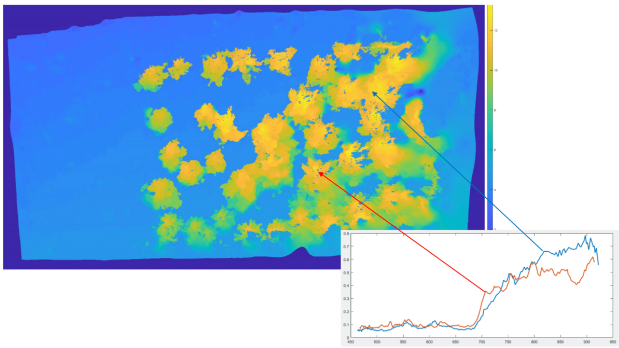

Fig. 6. A hyperspectral data cube created from multiple images of wooded area captured by a Glana hyperspectral sensor. In each spatial position, spectral information of multiple wavelengths is available.

Fig. 7. The scene with the airport control tower and the "crime scene" was reconstructed by Spotscale AB using a photogrammetry pipeline with 1735 images from a GoPro HERO7 Black camera.

cars is chosen for study, that is, how to infer the intention of the pedestrian at the kerb. The pedestrian problem itself is of significance in developing autonomous cars. Statistics show that pedestrians account for 65% of the fatalities out of the 1.17 million worldwide traffic related deaths. Furthermore, around 80% of the accidents with injuries in an urban traffic environment caused by a pedestrian error are attributable to misbehavior when crossing the street. A number of approaches have been proposed to tackle the problem. One of the promising solutions is to use the body and body gestures to infer pedestrian intention. Due to inherent subjectivity and ambiguity in human intentions, limited success in intention estimation has been achieved. A new way to approach the problem of human intention is explored in our project: an interactional account for intention. In this account, human intention is no longer an objectively measurable unit with well-defined category classes, but rather a dynamic processing, which is subjective and inherently ambiguous. It is argued that the phenomenological works of Heidegger and Merleau-Ponty could offer us a framework to characterize human intention. Within the framework, instead of treating our body as an objective object as commonly done in the engineering way, we see our body as a phenomenological body through which we could experience and explore the world. In the project we have

Fig. 8. 3D model from the Västervik measurement campaign created using NFC's lidar to model a scenario that illustrates a crime scene.

demonstrated how the phenomenological body could be used to constitute a bodily space for actions and how the body was used to coordinate the captured information to reconstruct a phenomenological scene: first-person view of the world. It is impossible to construct such first-person views without the high-resolution, precise n-dimensional models developed in this project. It is the first-person view that provides us a platform where the interaction between people and environment could be characterized. Furthermore, we have demonstrated how to define walkability, a street affordance, and how to compute it from the first person view used for inferring long-term pedestrian intention, see [6].

The Västervik campaign also included a rich set of other measurements, for example to demonstrate modelling of urban areas, forests, underground- and underwater scenes. Creating a precise and accurate n-dimensional model requires methods for automating data collection and processing to handle the large amount of data and to speed up the work, from the measured data to the finished n-dimensional model. An example of fast and automated processes is Vricon AB's visualization platform for creating 3D models from satellite data. Moreover, the time between an event and the measurement is especially important if you for example want to model a crime scene before it may change too much of weather and activity on the site or to support rescue operations in the case of wildfires.

IV. Conclusion

In summary, the project demonstrates the use of n-D as an important component and basis for image analysis and visualization. We give some examples of how the Västervik data can be used for models beyond today's 3D techniques by intelligent n-dimensional models of large outdoor scenarios. The concept provides unique information derived from the multidimensional sensor data, and any results re-used to retrain databases, update algorithms or in other ways exploit experiences and knowledge. A strength of the platform project is to network in a way that honors synergy effects through collaboration in a unique laboratory environment, with several different imaging sensors. The platform is thus formed by a rich network of algorithm developers, sensor providers and end users that contributes with sensors, algorithms and visualization techniques. In this environment the different stakeholders exchange challenges, hardware, and algorithms, new applications can be developed using knowledge from different actors. This in turn have created access to an environment that the individual partners cannot create on their own without effort and investments.

A key issue has been to make the data-sets accessible for the project participants. Sensor data is, for example, disseminated to researchers in computer vision and visualization. Results from the platform project are expected to further create value by supporting the development of new services and products within the business portfolio of participating companies. The platform project thus support regional and national growth by communicating knowledge about different sensor types and the latest technologies within computer vision and visualization.

A new project within Visual Sweden is also started, *Smart Twin for Forest Environment*, to exploit how n-D models can support the combination of environmental friendly and cost efficient forestry. Other applications are e.g within surveillance, command & control, rescue operations, environmental monitoring, autonomous vehicles and agriculture.

Acknowledgment

The project *Intelligent n-dimensional modeling by multidimensional sensor informatics for computer vision and visualization (n-D modeling)* is funded by Visual Sweden, through Sweden's Innovation Agency (Vinnova), and includes; Astacus, Mainbase, KTH, RISE SICS East, RISE Interactive, Interspectral, Spotscale, Vricon, Voysys, Termisk systemteknik, Glana, Deep Forestry, CVL and ITN at Linköping University, Swedish Defence Research Agency (FOI), Swedish Police including NFC, Smart Agritech Solution, Blink Services and Deep Vision.

References

[1] H. Ovrén and P.-E. Forssén, "Trajectory representation and landmark projection for continuous-time structure from motion," *International Journal of Robotics Research*, vol. 38, no. 6, pp. 686–701, May 2019.

[2] C. M. T. Anderson, F. Järemo, and P.-E. Forssén, "Assessing losses for point set registration," *Robotics and Automation Letters*, February 2020.

[3] J. J. Gibson, *The Ecological Approach to Visual Perception*. Houghton Mifflin Harcourt (HMH), Boston, 1979.

[4] F. J. Lawin and P.-E. Forssén, "Registration loss learning for deep probabilistic point set registration," in *International Virtual Conference on 3D Vision (3DV 2020)*, November 2020.

[5] M. Merleau-Ponty, *Phenomenology of Perception*. Routledge & Kegan Paul, London, 1962.

[6] H. Li and L. Klasén, "Human intention prediction and its application in forensics," in *FG2020 Tutorial: 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG2020)*, November 2020.