VISUAL SWEDEN

RESULTATRAPPORT

Intelligent n-dimensional modeling by multidimensional sensor informatics for computer vision and visualization

Projektnummer 2015-07051

Martin Holmberg & Lena Klasén

Innehållsförteckning

1. Summary

This report contains a description of results from the platform project “Intelligent n-dimensional modeling by multidimensional sensor informatics for computer vision and visualization”, including goals, participants, and results. The project is detailed in the Final Project Report.

The number of participants was very high, resulting both in challenges and possibilities. The main advantage that we managed to achieve in the project was to create an environment, a platform, where the partners shared experiences, learned from each other, creating new commercial opportunities and new collaborations. On the other hand, it was difficult to maintain a clear common vision for all partners coming with different needs and background.

Scientifically speaking, advances have been made in algorithm development for data registration and visualization, resulting both in scientific publications and new commercial services. Detailed scientific results are presented in a number of papers and therefore omitted in this report. Some results are already commercial products, while some will be brought forward for further studies in the new Visual Sweden project “Smart twins for forest environments”.

2. Background, purpose, and goal

In the project, a platform model was developed to visualize an area in several spectral wavelength ranges, all spatial dimensions, and several other dimensions or sensor modalities, such as radar (n-D). Data was collected with several different sensor types. With the data collected a database was created. The platform project enabled collaboration between algorithm developers and end users, such as companies and more applied research projects. The platform project thus further developed today's high-resolution, precise 3D models into unique and intelligent n-dimensional models for analysis and visualization of complex environments and events and thus provided means for new products and services for the participating industries whilst feeding back ideas for further research topics to academia.

The purpose of the project was to enable collaboration between algorithm developers and end users, such as companies and more applied research projects. This was done by facilitation of method refinement and adaptation to specific needs, sensor types and applications.

The goal was to provide the partners in the project with enough knowledge to be able to bring the results to the market within a short time frame. The project thus further developed today's high-resolution, precise 3D models into unique and intelligent n-dimensional models for analysis and visualization of complex environments and events and thus provided means for new products and services for the participating industries whilst feeding back ideas for further research topics to academia.

3. Organisation and conduction

The following organizations were partners in the project, and their roles are described below:

Vricon: 14 satellite images taken on Monday 10:03–10:35 during the measurement campaign from three satellites of which two with 46 cm resolution and one with 31 cm resolution. Based on these, three 3-d models were uploaded to our database: one based only on new images, one based on older images, and one based on all images. Also images collected from an UAV.

Spotscale: GoPro images and Osmo Pocket images have been collected. Point clouds and textured 3-d models were produced. Spotscale also applied to the National Land Survey (Lantmäteriverket) for the dissemination of all data produced in the project.

Swedish Police / NFC: A terrestrial laser scan of the control tower and the car scene (a staged crime scene, featuring a car, a mannequin and various items of evidence). The result was a point cloud with colored points. Also, a handheld Mantis Vision scanner (based on structured light and photogrammetry) was used to record details of the car scene. These data sets can be fused and visualized jointly by registration.

LiU / CVL: Occam Omni Stereo camera, a 360-degree stereo camera with a total of ten cameras, five per ring was used to collect data at the urban site. In addition, GoPro recordings were made at the urban site, both during daytime, and at dusk. A spinning Lidar (Ouster OS-1 64) was carried on a tripod and used to record the city square, as well as parts of the forest and the control tower at the airport. The Lidar data were registered using the algorithm developed in the project.

Astacus: Ground penetrating radar (GPR) and a terrestrial laser scanner (Leica P40). The laser scanner point clouds are colored, and georeferenced at the cm level. GPR recordings were made on urban streets, and around the airport control tower. They show subsurface features, such as e.g. wires and pipes. There are a lot of objects underground in the city, almost nothing at the airport.

RISE SICS East: Flights with Glana's hyperspectral camera and aerial photos from DJI Mavic. These are available, but the aerial photos were not further analyzed.

Swedish Defence Research Agency (FOI): UAV flights with a custom-built 3D-UAV equipped with a velodyne LiDAR. Data were collected during several flights in the urban site, both over the control tower and a rural scene in the end of the runway including a forest edge and a car. In the morning of the second measurement day, data was also collected at the urban site around the city square and pier. Ground reference points and other reference points on buildings and other structures were collected with high spatial coverage in both the urban and rural area for joint registration of the sensor data from the different partners. They were documented with visual reference and the coordinates were provided in different global reference systems to fit the different sensors. These georeferenced points were collected using a total station and RTK GPS. In addition, coordination of the measurement campaign was performed by dedicated personnel from FOI.

RISE Interactive: Produced visualizations of the data from the other partners.

Glana: A compact hyperspectral camera, with an optical filter which produce different transmission for different wavelengths. Image processing for merging the channels / pixels and 3D info, primarily in the forest in order to discover stressed trees (for example by bark beetles). Hyperspectral data was also collected over the airport terminal and the simulated crime scene.

Interspectral: Was supposed to visualize data and create interaction with data. They did not come back with info on the progress.

Mainbase: Carried sensors in their Apid in the Västervik campaign. Also reported gamma sensor measurements made in Västergötland. Four different types of sensors were mounted on the Apid.

Termisk Systemteknik: Mainbase flew one of their IR cameras, an uncooled FLIR LWIR. Two scenarios around the tower were recorded. A 3D model of the tower was produced.

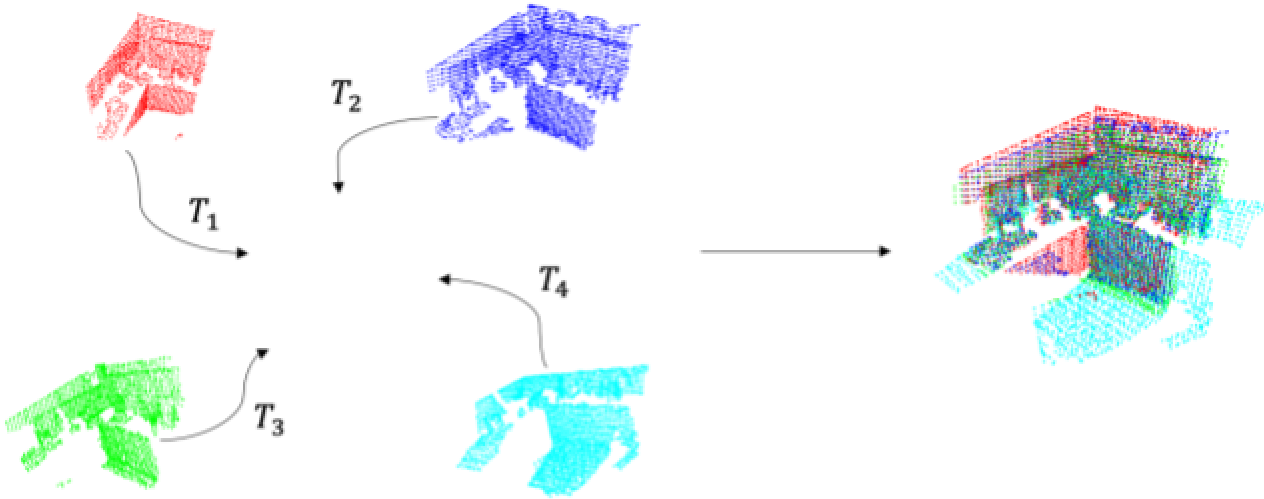
The datasets produced in the project are shared with the project partners using a file server hosted by Spotscale.

In the project, physical meetings were held 3-4 times per year until the pandemic. The meetings were held at the premises of one of the partners, giving room both for presentation of the host partner, and for a project meeting. One measurement campaign over two days was held in Västervik in May 2019, which gathered almost all partners. The day-to-day running of the project was handled by the project management group. Towards the end of the project, a data group was formed to handle data format, data storage and visualization. The project period was prolonged and a final digital workshop were held 24th September 2020. We presented the project at Visual Forum 30th September 2020. An open Visual Sweden seminar was held 25th February 2021.

4. Results

Work on point set registration

A fundamental component when fusing smaller 3D models into a large one is point-set registration (PSR). Individual point sets can come directly from Lidar sensors, or indirectly from image data, via structure from motion. The task in PSR is to find the unknown transformations that aligns point sets sampled from the same scene, see Figure below:

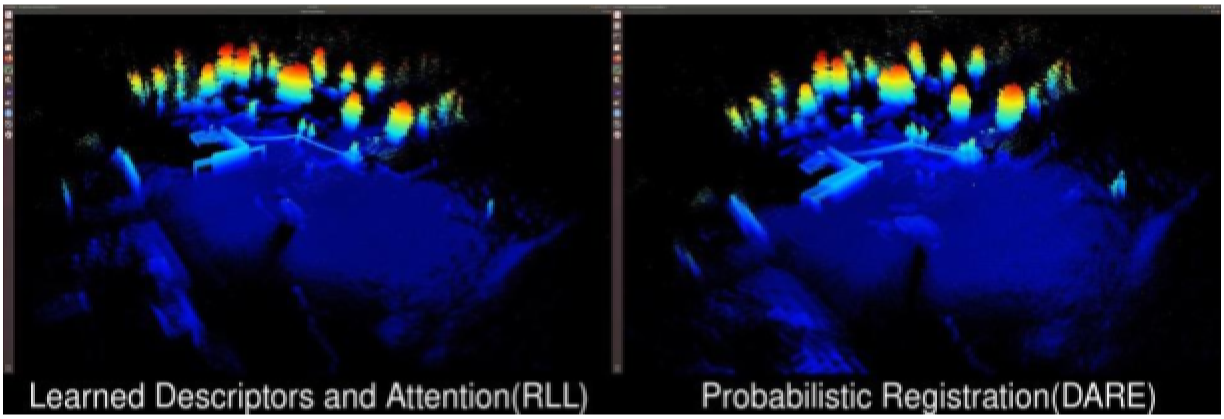


Point set registration. Left: four individual point sets, and sought transformations Ti. Right: the fused point set.

We have done an analysis of the principles of point set registration, and compared different choices of registration cost functions (losses). This was published in the journal *Robotics and Automation Letters*. The conclusions were:

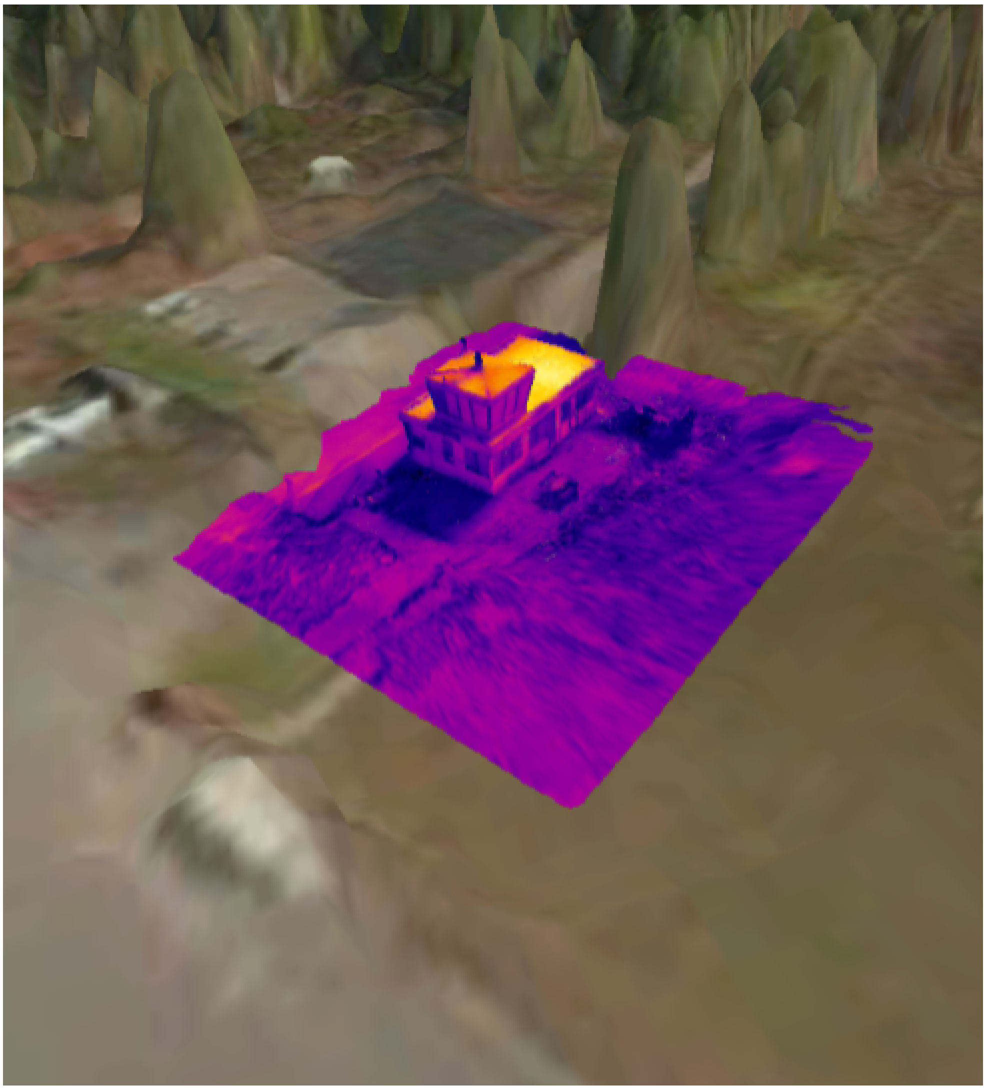
- Point based loss functions such as used in ICP are better than shape based losses far from the solution
- Kernel based loss functions are more robust
- Robustness is always improved when performing density-based weighting of the loss

We also integrated deep features into a probabilistic method for registration, and derived a fully differentiable registration pipeline. This allows the registration loss to be used for training the deep features (called RLL, registration loss learning), given examples of successful registrations. The deep feature descriptors and the point-wise attention weights were learned on a separate datatet, using a registration loss as supervision. A comparison is shown in the figure below. The learning framework was published at the *International conference on 3D Vision*. Attention weights have great potential to avoid incorrect registrations of repetitive structures, such as windows on large buildings, and trees. We intend to explore the full potential of the RLL framework in future projects, by tuning the descriptors and attention weights to specific scenarios, e.g. forests.



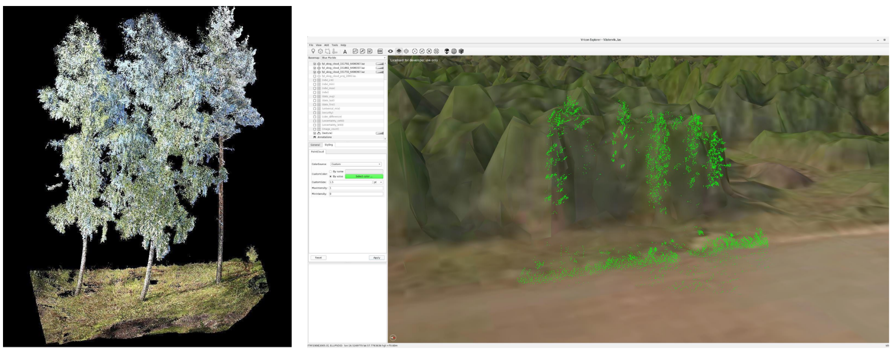
3D model of buildings and trees at Västervik airport. Left: Using the learned deep features. Right: using the plain probabilistic registration pipeline. As can be seen, the learned descriptors give more precise registration, and less drift.

One visualization was created in Unity, using data from Vricon, Spotscale and NFC. Further improvement of the model could have been made if there was a chosen use-case (e.g. forestry, archeological investigation, etc).



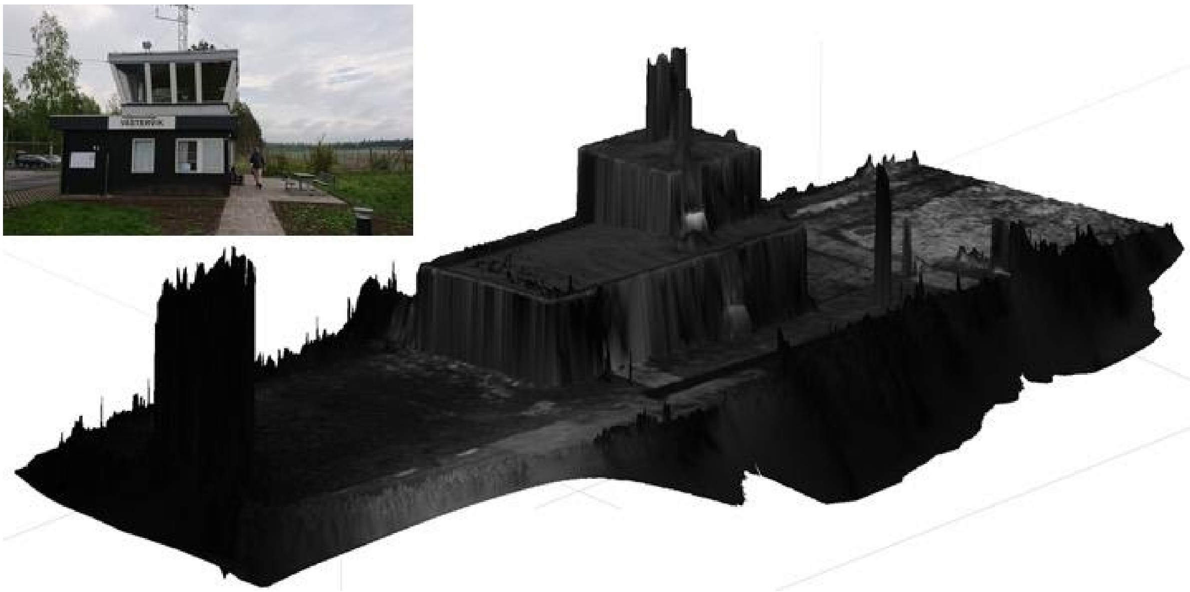
Reconstruction of flight tower at Västervik airport, using satellite data, visual data, and Lidar data.

A comparison was made of reconstruction of trees using satellite data from Vricon and Lidar data from Astacus. A comparison can be seen below. It was concluded that simultaneous registration would have been necessary to incorporate the data into one model with a high quality.



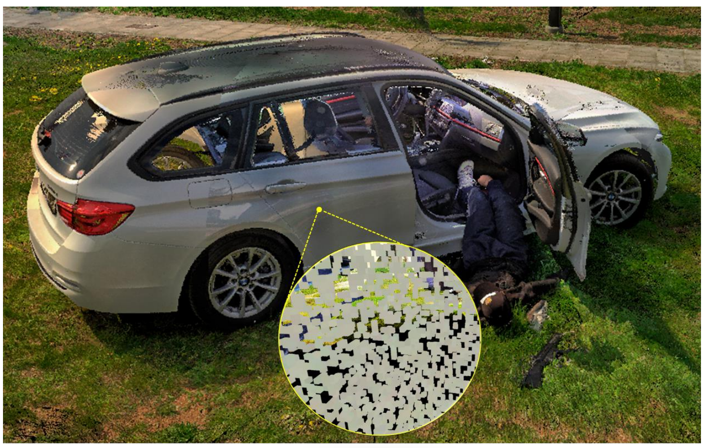
Comparison of reconstruction of trees (Lidar to the left, combination of satellite and Lidar to the right)

The hyperspectral reconstruction has been developed. Still to make full reconstruction (hyperspectral and 3D) of all collected data, but the software almost there! We have the entire processing chain, but still a few hacks are necessary.



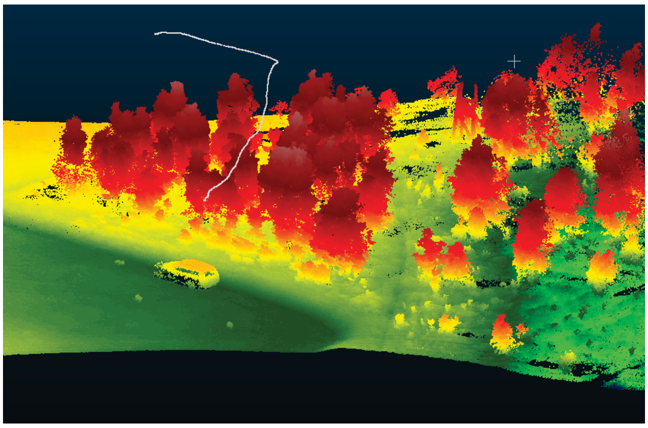
3D model of the airport control tower using images from the Glana hyperspectral camera.

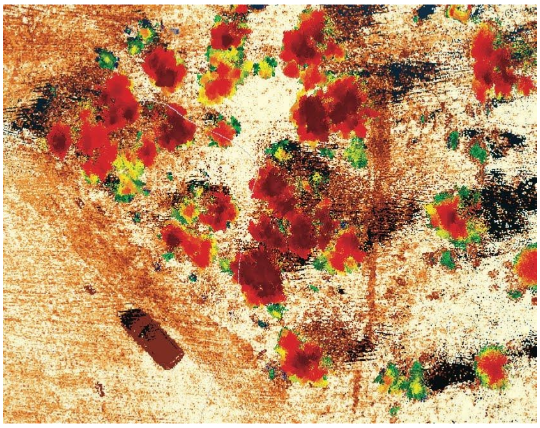
The “crime scene” was reconstructed with the purpose of creating a digital twin that could be used for forensic studies after an event. The reconstruction is shown below, with data coming from laserscanner FARO Focus M70. Viewing a 3D-model from a couple of meters make the model look very dense like surfaces but is very narrow placed points, a point cloud. Because of the very good point accuracy which is only a few millimeter makes 3D-models very useful to describe forensic scenarios. 3D-models like this can be used for further analysis such as bullet trajectories and reconstructions. Colorization of a point cloud is normally RGB while it is important for recognition showing the 3D-model in court.



Crime scene reconstruction.

Furthermore, the airport was visualized using 3-d laser, using a novel method for fast and accurate LiDAR semantic segmentation. Results are shown below.

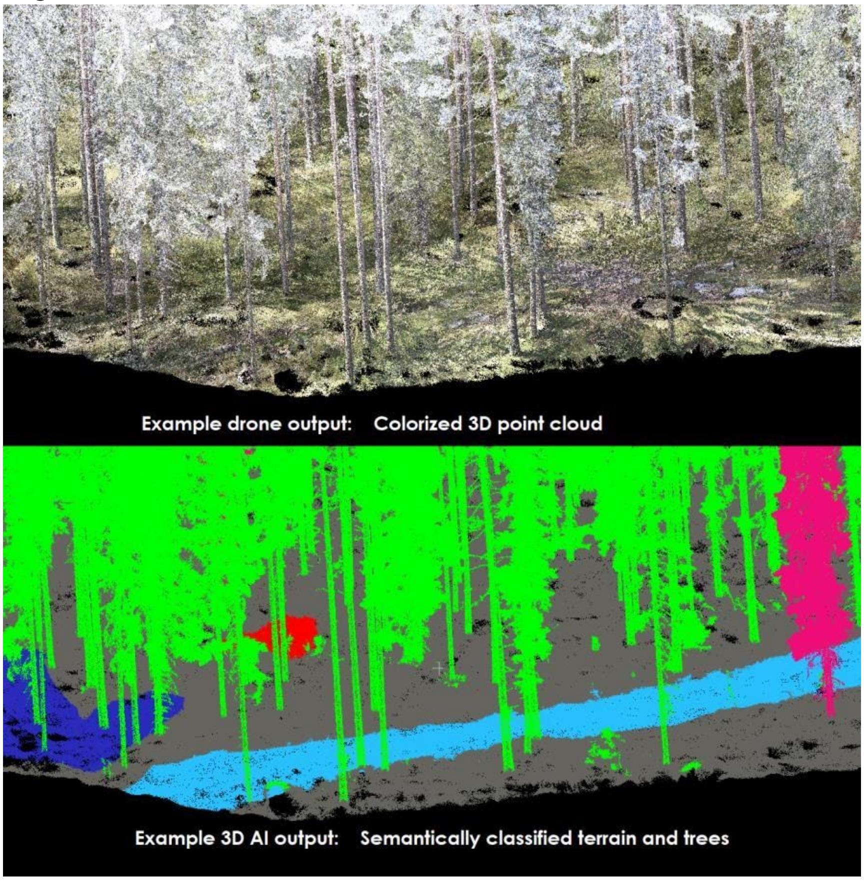




Point cloud registration

Images of the forest (below), and the Lidar reconstruction results (above).

The forest was reconstructed and classified using AI and point clouds from an RGB camera mounted on a drone. This made it possible to classify and visualize different features of vegetation and ground features.



Sensor data from drone (above), and classification of vegetation and ground features (below).

A landscape-oriented approach was developed to allow for allowing for inclusion of human behaviour in the model. We created the map with the height and slope. The slope is highly related to "gradients", a surface characteristic of an area. A machine learning approach has been developed to classify the environmental data into different qualitative categories. These categories have direct correspondence in the real world and can be observed by people through simple perception. The classified categories should get close to the categories employed by people when acting within the landscape.

5. Reflections and future work

As mentioned, a possibility and challenge included a clear common vision for all partners such as an application example. Thus, the workshop in December 2019 outlined a possible direction for a follow up project. During the final pase of the project, and a smaller "transition" project" sustainable forestry was choosen as target application. This lead to the new Visual Sweden project "Smart twins for forest environments". Moreover, a Formas application "Visual Forest" is submitted. CVL and Deep Forestry are in cooperation with SCA working in a Vinnova funded project that started in November 2020, to take an existing 3D AI algorithm SeMaFore and improve the software and prepare it for training with the

SCA data. These projects will bring the n-D concept further. The plan is also to use the n-D data in WASP PhD student courses and projects.

References

In the project, two films were produced describing the project and the results. Furthermore, a number of publications have been made:

- 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.
- C. M. T. Anderson, F. Järemo, and P.-E. Forssén, “Assessing losses for point set registration,” Robotics and Automation Letters, February 2020.
- Haibo Li and Lena Klasén, “Inferring Pedestrian Intention by Using the Human Body”, Late-Breaking Paper, 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG2020), 2020.
- Haibo Li and Lena Klasén, “Human Intention Prediction and Its Application in Forensics”, FG2020 Tutorial: 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG2020), 2020.
- Lena Klasen, Amanda Berg, Per-Erik Forssén and Haibo Li, “Beyond 3D Imaging and Visualization”, 38th Annual Swedish Symposium on Image Analysis (SSAB), 2020/2021.
- Haibo Li and Lena Klasen, “Predicting Human Intentions: A Phenomenological Approach”. Manuscript, 2020.
- A. C. M. Tavares, F. J. Lawin and P.-E. Forssén, "Assessing Losses for Point set Registration," in IEEE Robotics and Automation Letters, vol. 5, no. 2, pp. 3360-3367, April 2020, doi: 10.1109/LRA.2020.2976307.
- F. J. Lawin and P.-E. Forssén,” Registration Loss Learning for Deep Probabilistic Point Set Alignment”. International conference on 3D Vision (3DV), Nov 2020
- Abdelrahman Eldesokey, Michael Felsberg, Karl Holmquist, Mikael Persson, "Uncertainty-Aware CNNs for Depth Completion: Uncertainty from Beginning to End", 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Conference on Computer Vision and Pattern Recognition (CVPR), 12011-12020, 2020.