Collaboratec project

CIVIT Technologies: **White Paper**

The Collaboratec project builds a collaboration model through which CIVIT research infrastructure and the know-how of new immersive technologies can be used effectively for companies' development needs.

This white paper presents the results of technology profiling that was carried out as a part of the project. The aim was to identify the features, possibilities and limitations of equipment such as state-of-the-art cameras, range sensors and other immersive gear in CIVIT. The findings reported in this document can increase the awareness of CIVIT's technologies and help to establish new academia-industry collaboration.

Created by the Collaboratec project team
www.tuni.fi/en/research/collaboratec

31.8.2023

Leverage from the EU 2014–2020
European Union
European Regional Development Fund

Table of Contents

Introduction 2
Technologies 2
Volumetric capture 3
Depth and hyperspectral sensors 7
Motion capture 14
Conclusions 18

Introduction

The Centre for Immersive Visual Technologies (CIVIT) is a national research infrastructure, funded by the Academy of Finland and Tampere University. It provides state-of-the-art equipment, annotated data and expertise in the sectors of 3D visual scene sensing, processing of visual data and immersive displays. The CIVIT portfolio of services targets three groups of users: academic research communities, talented students, and innovative companies with related services to enable cutting-edge research in the field of immersive visual technologies.

To strengthen the link between academia and industry, CIVIT has developed a collaboration model through which the CIVIT research infrastructure and the know-how of new immersive technologies can be used more effectively for companies' development needs. The work was done in project "Collaboration model for using immersive visual technologies in work machine and creative digital industries" (Collaboratec), funded by the European Regional Development Fund.

One part of the development of the collaboration model was to profile the technologies available in CIVIT and report the results in white paper format. The aim was to identify the possibilities and limitations of CIVIT's equipment to increase awareness of CIVIT's capabilities as well as to enable quicker experimentation and collaboration projects based on company needs.

This white paper describes the findings of the technology profiling that was linked with other tasks in the project such as developing software and hardware libraries, creating datasets and implementing proof of concepts (PoCs). These activities contributed to the profiling by providing information on how to best integrate and utilize the equipment in different use cases. The technologies are presented in dedicated sections, which provide brief technical overviews, describe the use cases, introduce the insights gained and explore open questions and potential research directions that could benefit future collaboration with companies.

Technologies

The aim of modern visual technologies is to convincingly reproduce reality and relay information in a natural way. During the project period, the profiling activities were primarily related to technologies intended for capturing visual information and displaying the captured content. In addition, the selected technologies had clear potential to be utilized in collaboration projects with companies. For these reasons, this white paper focuses on a subset of the technologies and equipment available in CIVIT. Table 1 provides a general overview of them.

Table 1: The different types of technologies covered in the white paper.

Sensors for understanding the environment	Displays for visualizing information
• Human appearance and actions • How the world or camera moves • Objects in the world and their measurements and pose	• Presenting 3D video content on XR headsets and 2D displays

The focus was on profiling sensors which capture different information from the environment such as colour and depth of objects, motion of the sensor and wavelength of light coming from the object. The other part of the profiling was related to exploring how to integrate

information captured and processed from these sensors into different display devices such as XR headsets, AR enabled phones as well as traditional 2D screens.

Volumetric capture

Overview

Volumetric capture is a technology that enables the creation of realistic 3D representations of real-world objects, people, or environments. CIVIT's volumetric capture studio (see Figure 1), developed by Mantis Vision, has 32 camera units arranged in a cylindrical configuration, spanning approximately 2 meters in diameter and 3 meters in height. Each camera unit has two RGB cameras and one structured light sensor. These camera units can capture the appearance and movement of a person from different angles simultaneously. The volumetric studio is equipped with a microphone to capture audio along with the visuals.

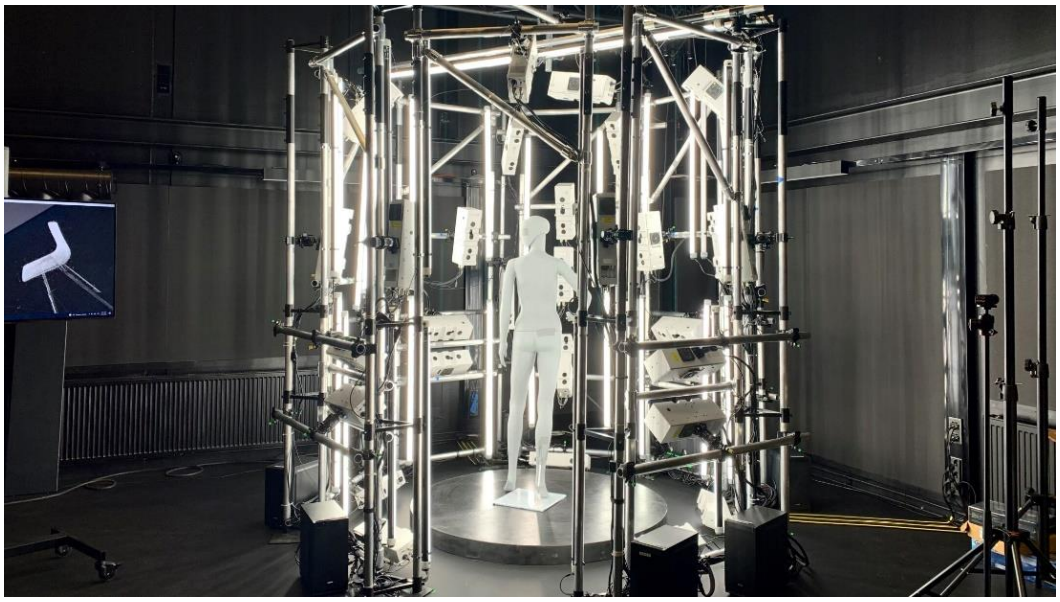


Figure 1: The volumetric capture studio.

The captured data consists of an extensive set of data points, which are used to create a dense point cloud and mesh that accurately represent the person's geometry, texture and spatial information. The volumetric video (in proprietary format) is captured at 25 fps. With the help of further software algorithms, the result is a seamless, high-fidelity 3D model that can be utilized in numerous applications such as virtual/augmented reality, teleconferencing, video games, and artistic performances. The volumetric video can be integrated into game engines and website code using Unity, Unreal Engine and JavaScript plugins. The volumetric capture can also be exported into standard 3D formats such as .obj, .glb and .fbx.

Use case: Violinist in VR

In the Collaboratec project, the volumetric capture studio was utilized in use cases related to the creative digital industries. In the first case, a solo concert piece composed for a Finnish virtuoso violinist Eriikka Maalismaa was captured in the studio (see Figure 2a). The recorded audio and visual data were used in a VR experience (see Figure 2b). In this experience, a person wearing a VR headset could hear the violinist playing in a forest setting. The concert piece was a combination of the sound of the violin and the forest, with the voices complementing each other.

(a) (b)

Figure 2: The violinist in the volumetric capture studio (a) and in the VR experience with a forest setting (b).

Use case: Immersive fashion

The second case explored the possibilities of using volumetric capture for digital fashion, with a particular focus on jewellery industry. The approach was to combine a volumetric capture of a person with a 3D model of the jewellery piece. We captured footage of an actor in the volumetric studio. The resulting 3D model can be used in creating a virtual try-on experience where a piece of jewellery such as a necklace or bracelet can be viewed on the avatar (see Figure 3). These volumetric recordings could be integrated into both 2D platforms such as browsers and mobile apps as well as XR headsets. The demos can be viewed online at https://devangini27.github.io/.

Figure 3: Screenshots of the immersive fashion demos.

Use case: Game and metaverse characters

In addition, an actor was scanned to create a game character for immersive games and an avatar for metaverse platforms. For this case, one frame was scanned with the volumetric capture studio and then rigged to create a skeleton of the avatar. Using the skeleton, the avatar could be animated to create various types of action sequences which would be suitable for game development. This PoC helped to generate recording recommendations for persons with the volumetric capture studio for rigging purposes. We identified the optimal pose as A-

pose with the hands in front of the chest as shown in Figure 4 (top left). Figure 4 also illustrates some of the poses which were generated using the rigged model and motion capture data.

Figure 4: Screenshots of an avatar created with a rigging software. The original capture pose is shown in top left. The other five examples show custom game actions created from the rigged skeleton.

Recording guidelines

The completed uses cases helped to profile the possibilities and limitations of the volumetric capture studio in CIVIT. The main findings were summarized in a list of recommendations that can be used as a basis when planning to capture new content with companies.

1. Wear brightly coloured clothes with texture. Black colour is challenging to capture and can result in artifacts in the 3D model.
2. Avoid eyeglasses. Due to the small dimensions and reflecting parts, glasses are challenging. Contact lenses are recommended.
3. Have hair tied. Long hair often introduces artifacts that can be challenging to remove.
4. Plan all movements to be done within an area of 1.6 meters in diameter. This is the size of the capture area. Movements performed outsize of this area will not be visible in the resulting model.
5. Use slow and medium speed movements. The studio works well with both slow and medium speed of movements. Actors can talk, use gestures, show facial expressions and these would be captured. Fast movements result in motion blur.

6. Have large accessories instead of small ones such as a ring. Accessories smaller than a wristwatch can be challenging to capture and model properly.
7. Avoid dark, thin and glass material objects. Capturing dark, thin and glass objects is not recommended.
8. Avoid objects with reflective material. Objects with materials such as metal may result in models with artifacts or holes.
9. Short captures are better than long ones. Capturing several short sequences (e.g., up to 30 seconds) accelerates the workflow in the studio compared to capturing a single sequence lasting for 2 minutes. Checking the quality of the capture and doing retakes is quicker with short captures.

These guidelines were utilized to create a non-commercial dataset in collaboration with Ultra Video Group (UVG) in Tampere University. Figure 5 illustrates some example images from this dataset. This dataset can be downloaded from https://ultravideo.fi/UVG-VPC/.

Figure 5: Examples of sequences from the UVG-VPC dataset

Open questions and research directions

There are several other commercial volumetric capture studios available, and it would be informative to compare the capture quality and postprocessing results with CIVIT's studio. This comparison would provide further understanding of the possible limitations and potential research directions such as improve postprocessing algorithms.

Further research directions include improving the current capturing ability of the studio as well as increasing the possibility of developing applications. The jewellery PoC as well as the VR orchestra case needed the volumetric capture studio to be able to scan small and reflective objects. Currently, these are limitations of the studio. It would be helpful to explore solutions such as including additional sensors into the studio to capture small and reflective objects.

At the moment, the captured data can only be streamed on local networks, and the volumetric video is compressed to be able to play it in real time on devices such as mobile phones and VR headsets. Future research is needed to explore improved methods to stream volumetric videos across the internet. Further methods for compressing the volumetric data structure without losing many details are also needed so that it would be possible to increase the details of the postprocessed 3D model.

References

- https://mantis-vision.com/

Depth and hyperspectral sensors

Overview

Depth sensors capture depth information of a scene. Solutions for this include Time of Flight (ToF) cameras, structured light sensors, stereo cameras and laser scanners. For estimating the depth, these sensors use different underlying principles such as time for light beam to return (ToF, laser), finding the pixel shift between two views (stereo cameras), and finding the changed reflectance of projected pattern (structured light sensor). The volumetric capture studio uses stereo and structured light sensors to estimate depth.

Hyperspectral cameras can measure wavelengths beyond the visible range. This can be used for material detection.

Use case: Immersive fashion

A process was developed to scan jewellery pieces with depth sensors, post-process them and display the final scans with an accessible viewer. The captured pieces can be used, for example, in developing immersive experiences that utilize content from volumetric captures. Jewellery pieces provided by the collaborating companies (see Figure 6) and test objects in CIVIT were scanned. The jewellery pieces were either made of wood or metal.

(a) (b)

Figure 6: Sample jewellery pieces used for profiling scanners. Jules & Beryl's "Aloe" (https://www.meirarauta.fi/fi/gallery.html) was made from copper and had iodized parts in green and reflective parts in orange (a). Birch plywood bracelet from Voaki (b) (https://www.instagram.com/p/CDRc-oEsufX/?hl=en).

The depth sensors tested in jewellery scanning are listed in Table 2 and shown in Figure 7.

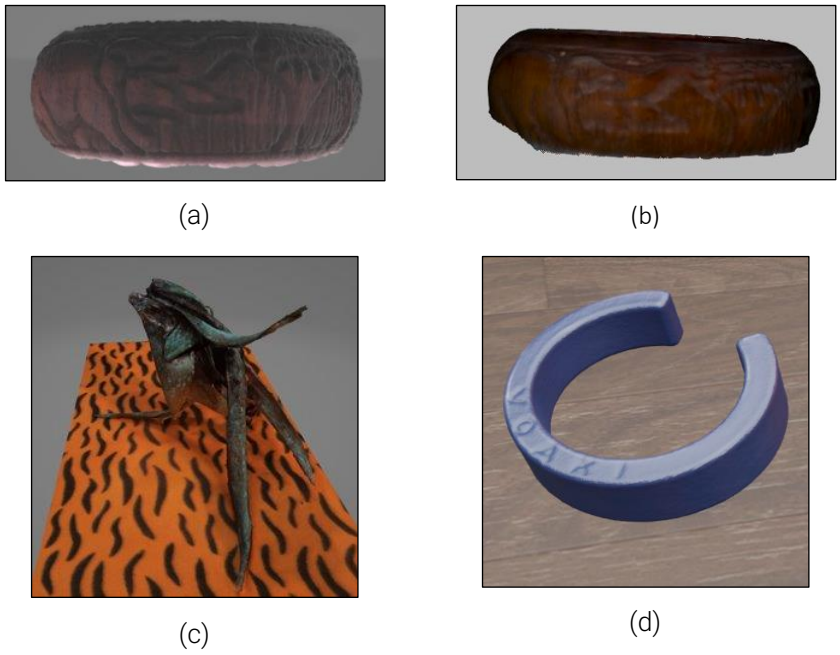
Table 2: Scanner specifications.

Scanner	Object size	Scan accuracy	3D resolution
Matter and Form	Maximum height: 25 cm Maximum diameter: 18 cm	Within ± 0.1mm	Not mentioned
Artec Eva	Minimum around 10 cm	<= 0.1 mm	<= 0.2 mm
Artec Leo	Up to 160,000 cm^3	<= 0.1 mm	<= 0.2 mm
iPhone ToF camera	Depends on the software used	Depends on the software used	Depends on the software used

(a) (b) (c)

Figure 7: Scanners profiled included Artec Leo (a), Artec Eva (b) and Matter and Form (c).

Laser scanners are ideal for jewellery pieces that are not reflective. Out of the three laser scanners, Matter and Form had the best texture resolution. On the other hand, it also had a limited scan area, so larger jewellery pieces were scanned with Artec Leo. Artec Eva was found to be unsuitable for scanning jewellery pieces due to its large scan area. iPhone scanner was able to scan reflective jewellery pieces better than other scanners though the resulting model had some artifacts such as missing small ends of jewellery. Artec Leo is not a part of the CIVIT equipment catalogue, but it can be accessed when needed. Example scans are shown in Figure 8.



(a) (b)

(c) (d)

Figure 8: Scans of different jewellery pieces. Wooden bracelet scanned with Matter and Form (a), wooden bracelet scanned with Leo scanner (b), Aleo piece scanned with iPhone scanner (c) and wooden bracelet scanned with Artec Leo.

A demo of an online viewer was developed using opensource front-end libraries so that the scanned jewellery pieces could be viewed on a web browser. The online viewer (see Figure 9) allows visitors to pan, scale and rotate the displayed model in 3D. When opened on a mobile device, the user can see the jewellery piece in augmented reality. The demos can be viewed at https://devangini27.github.io/.

Figure 9: Screenshot of a scan shown with the online viewer.

Open questions and research directions

The profiling of different depth sensors for scanning different jewellery pieces revealed the limitations of current hardware for this use case. Many of the tested jewellery pieces and objects had reflective parts that are difficult to scan properly. The current approach to circumvent this issue is to use 3D scanning sprays on the reflective parts. However, using these chemicals may not be possible with hand-crafted jewellery pieces with delicate parts and different metals. Future research could be carried out to scan such reflective materials by combining scans from different sensors.

References

- https://matterandform.net/
- https://www.artec3d.com/portable-3d-scanners/artec-eva
- https://www.artec3d.com/portable-3d-scanners/artec-leo

Use case: Camera tracking for virtual effects

Camera tracking is an algorithm used for virtual effects (VFX) technique called match moving. Match moving is a VFX creation process, in which virtual objects are augmented into live video with appropriate position, orientation, and scale to match the background objects of the video. In this use case, different depth sensors were tested for their suitability for camera tracking algorithm. A total of six sensors were profiled: Kinect Azure, Intel RealSense SR300, Intel RealSense D435, ZED Camera, PMD Pico flexx and Basler ToF (see details in Table 3). Out of these sensors SR300 was rejected early on due to its limited range.

Table 3: Depth sensor specifications.

Sensor name	Range	Resolution	Field of view	Frame rate	IMU sensors
Kinect Azure	0.5–5.46 m	1024 x 1024 px	120° x 120°	30	Yes, 208 Hz
ZED Camera	0.2–20 m	1920 × 1080	110° x 70°	100	Yes, 400 Hz
Intel RealSense SR300	0.3–2 m	640 x 480 px	73° x 59°	30	No
Intel RealSense D435	0.3–3 m	1280 × 720 px	87° × 58°	90	No
PMD Pico flexx	0.1–4 m	224 x 171 px	62° x 45°	45	No
Basler ToF	0–13 m	640 x 480 px	57° x 43°	20	No

Drawbacks of using only depth information for tracking are the reliance on a powerful GPU to perform fast calculations and susceptibility to being affected by moving objects. Kinect Azure and ZED camera provide depth information as well as IMU (accelerometer and gyroscope) sensor information which can be used to increase the frame rate of camera tracking

algorithms. Also, IMU makes it possible to remove the effect of moving objects in the scene. Out of the Kinect Azure and ZED camera (see Figure 10a and Figure 10b), ZED was selected for developing a software library due to its longer range and additional polariser for dealing with different lighting conditions.

Figure 10: Depth sensors profiled in the camera tracking use case. Kinect Azure (a), ZED Camera (b), Intel RealSense SR300 (c), Intel RealSense D435 (d), DMD Pico flexx (e) and Basler ToF (f).

For a PoC, a user interface (see Figure 11) for integrating the rotation and translation values from camera tracking was developed in Unity with shaders. A rabbit was placed onto the scene initially and the camera tracking algorithm helped to move the rabbit relative to the background. This created an illusion where the rabbit stayed in its original position (for example, next to the chair as in the example figure) even if the camera was moving.

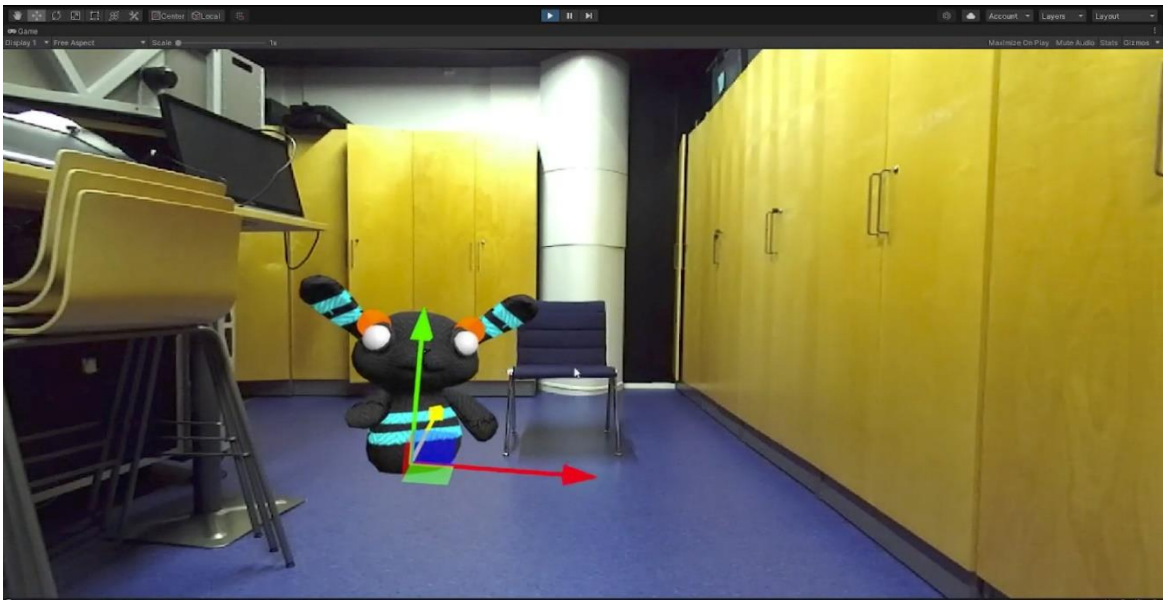


Figure 11: Unity based user interface for the camera tracking algorithm.

Open questions and research directions

The current PoC implementation was not able to segment people out properly. This feature would allow people in the video to interact with the virtual objects. Future research could investigate how to properly segment people in the scene and occlude the virtual object. Another limitation of the current the camera tracking algorithm is changes in environmental lighting. Future research could investigate other sensors that could be integrated into the match moving system to provide this feature. Also, further work could be done to benchmark the camera tracking algorithm with motion capture to determine its accuracy in different conditions such as size of space, moving objects, empty space, etc.

References

- https://learn.microsoft.com/en-us/azure/kinect-dk/hardware-specification
- https://www.stereolabs.com/zed-2/
- https://www.intel.com/content/www/us/en/products/sku/92329/intel-realsense-camera-sr300/specifications.html
- https://www.intelrealsense.com/depth-camera-d435/
- https://3d.pmdtec.com/en/3d-cameras/flexx2

Use case: Work target detection and measurement for the work machine industry

This PoC and related dataset targets the heavy work machine domain and addresses their need to detect and measure objects for safety and autonomous applications. Heavy machines such as cranes and forestry machines operate in environments which are not optimal for machine vision. The terrain (occlusion by environment or moving objects) and lighting (brightly lit, dark, shadows and light direction) are constantly changing (see Figure 12). Cameras must be placed on the machine which introduces shaking and occlusions from the machine itself.

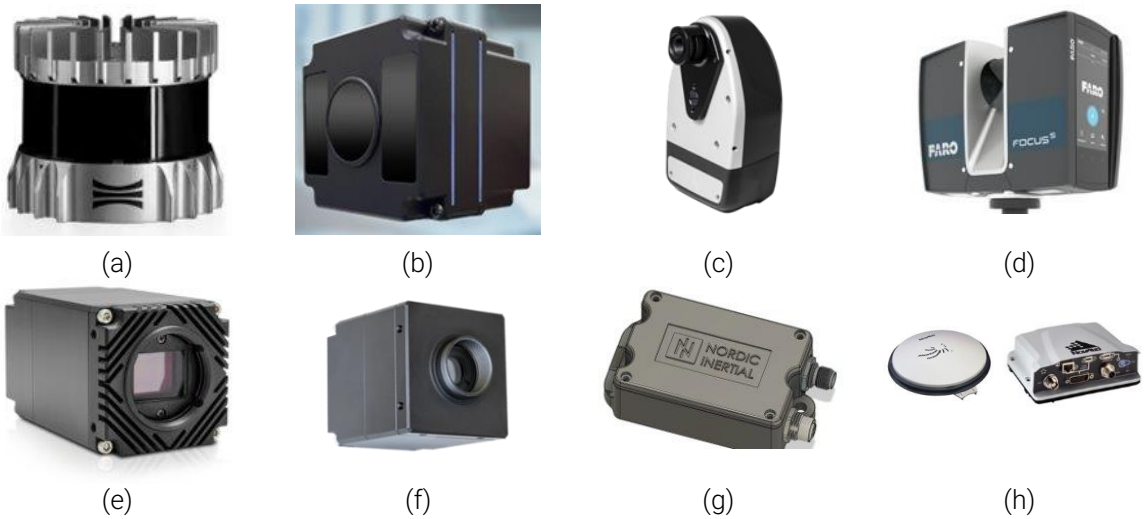
(a)

(b)

(c)

Figure 12: Challenging conditions in work machine settings include night working conditions (a), occlusion caused by leaves (b) and shadows on objects (c).

The aim was to create an example dataset and profile new sensors (stereo cameras, Time of Flight, short wave infrared camera and GNSS+IMU) that were placed on a wheel loader. Continuous data was collected while the vehicle was moving and handling target objects. The included sensors are shown in Figure 13.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 13: Sensors profiled in the dataset collection included Ouster LIDAR OS0 (a), OToBrite oToCAM500 ToF (b), Senop Rikola HSC-2 (c), FARO Laser Scanner (d), Lucid Vision IMX990 SWIR (e), Lucid Vision IMX490 (f), Nordic Inertial IMU (g) and Novatel GNSS (h).

The FARO Laser scanner and Novatel GNSS were meant primarily for collecting the ground truth data of the test environment. The environment was scanned from multiple locations with the FARO Laser scanner and combined into a single point cloud file. After post process, Novatel GNSS provided cm level of accuracy for localization in 1 fps. In-between frames are accurately interpolated with help of the KVH1750 IMU unit.

Table 4: Specifications of the sensors used in dataset collection.

Sensor name	Resolution	Measuring range	Sampling rate
Ouster LIDAR OS0	128 channels 90 vertical FoV 5.2M points per second	up to 100 m	20 Hz
OToBrite oToCAM500 ToF	640 x 480 62º x 47º	0.3 m ~ 7.2 m	15 fps
Senop Rikola HSC-2	1024 x 1024 pixels 36.8° x 36.8°	6-18 nm 1000 spectral bands	149 fps
FARO Laser scanner	360° X 300°	130m	5-45 fps
Lucid Vision IMX990 SWIR	1280 x 1024 pixels	400 nm to 1700 nm	84.9 fps
Lucid Vision IMX490	5.4 MP, 2880 x 1860 pixels	-	20.8 fps
Novatel GNSS		~1 cm accuracy	1 Hz
Nordic Inertial IMU			200 Hz

Most of the sensors supported the precise time protocol (PTP). Synchronization of the system was implemented with this. GNSS was the master clock in the system. A PC worked as a boundary clock, receiving PTP time from the GNSS and distributing it forward to the Ouster and Lucid Vision sensors. This way each image from these sensors had synchronized timestamps. IMU and ToF camera did not support PTP so those were saved with timestamp from the PC instead. As the PC was in PTP time, these frames were synchronized as well. This introduced a small delay as the timestamps did not represent the exact time when the images

were taken. However, by visually comparing the data from different sensors, this delay was eliminated by applying approximated offset in the post process.

All the sensors were attached to a custom sensor rig. This was to keep their relative positions consistent and to easily calculate certain measures between the sensors. The rig was constructed in such a way that it was easy to mount on different work machines. For this dataset collection, it was mounted on an Avant wheel loader (see Figure 14a). The Avant was developed and maintained by the Autonomous Mobile Machine group lead by Professor Reza Ghabcheloo.

(a) (b)

Figure 14: Sensor rig mounted on an Avant wheel loader (a). Processed data from the FARO Laser scanner showing the test environment.

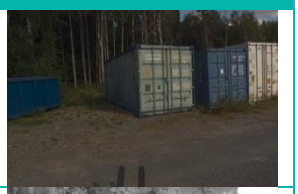
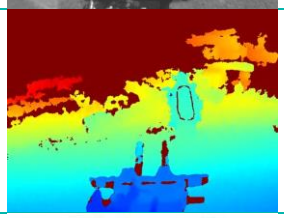
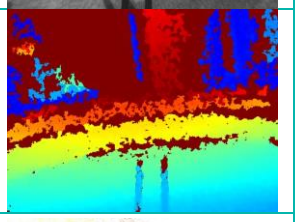
The test environment (see Figure 14b) consisted of three different work targets that were containers, logs and boxes (see Figure 15).

(a) (b) (c)

Figure 15: The work targets used in dataset collection: containers (a), logs (b) and boxes (c).

The next step was to post-process the collected data. For easy usage and reading, each frame from the sensors was converted to .png and scans to .ply format. GNSS and IMU data was saved in .csv format. Some samples from the collected open dataset are shown below in Table 5.

Table 5: Example data with different sensor types for two work targets.

Sensor type	Logs	Container
RGB image from Lucid Vision IMX490		
Lucid Vision IMX990 SWIR		
Depth image from OToBrite oToCAM500 ToF		
Point cloud from Ouster LIDAR OS0		
Hyperspectral image from Senop Rikola HSC-2		

Open questions and research directions

This multimodal dataset helps to overcome certain limitations of colour cameras. Using information from colour cameras is challenging when it comes to dark lighting conditions, changing light directions, weather conditions and occlusion. Depth sensors such as LIDAR and ToF work in dark environments and are insensitive to changes in illumination and sun direction changes. However, depth sensors sense the depth of objects that appear first in the scene and hence occluded objects cannot be sensed through depth sensors. Hyperspectral cameras provide information related to the materials and are insensitive to illumination, occlusion and work in different weather conditions. Using a combination of different modalities help to overcome these challenging conditions.

Motion capture

Overview

Motion capture (MoCap) is the process to capture an object or human movement. Notoriously, it has long been widely used as source of animations in film and game productions. Other use cases include sports science, ergonomics and XR applications. The traditional and most common MoCap technologies include inertial and optical systems.

Inertial systems are generally meant for human body capture. The system consists of IMUs attached to the body joints and bones. Data from these units is either recorded inside the suit or wirelessly transmitted to the host PC, which processes the information and outputs the body poses. Advantages of this system include portability and cheaper price compared to optical systems. The biggest downside is that the positions of the IMU units are not tied to anything, introducing drift of the position and pose over time. This makes the system as such unsuitable for precise tracking applications. This problem can be avoided by using some other tracking solution for localization alongside inertial suit.

Optical systems consist of multiple cameras spread around the capture volume that sees individual markers on the scene. These markers can be attached to the human body to create a skeleton or some other objects in the capture volume to create rigid bodies. The exact pose and 3D position of these bodies and objects is triangulated and calculated from the calibrated video streams. And unlike in inertial systems, here positions and poses are tied to the common ground, providing accurate positional data in relation to each other. This also guarantees that there is no temporal drift.

CIVIT has Xsens MVN link which is industry-leading inertial MoCap capture suit. The suit comes in a small portable package and is fast to set up and calibrate on site. This makes it most suitable for capturing individual animations of actors for games and film for example. Live applications in Unreal and Unity are also possible.

OptiTrack on the other hand is a state-of-the-art optical motion capture system. CIVIT studio has a fixed installation of OptiTrack Prime camera. In addition, there is a set of USB Flex 13 cameras for smaller temporary setups. In optimal conditions, OptiTrack can provide sub-mm accuracy without drift, and it is useful in applications where precise positional ground truth is required.

Use cases: 6DOF object tracking dataset

MoCap can provide, for example, an exact pose of a speaker and microphone, camera and target or swarm of robots or drones navigating in environment while avoiding each other. All previous examples are from previous use cases. Moreover, the system could be utilized in entertainment applications such as on-camera VFX, live productions and VR experiences. The starting point is to calibrate the exact pose of camera's optical center (or VR pivot point) in respect to the MoCap system's coordinate frame. Figure 16 shows the camera and different coordinate frames inside the tracking volume. Marker attached on top of the camera is the point the MoCap system reports as a camera position. The real point of interest however is the camera's optical center below it.

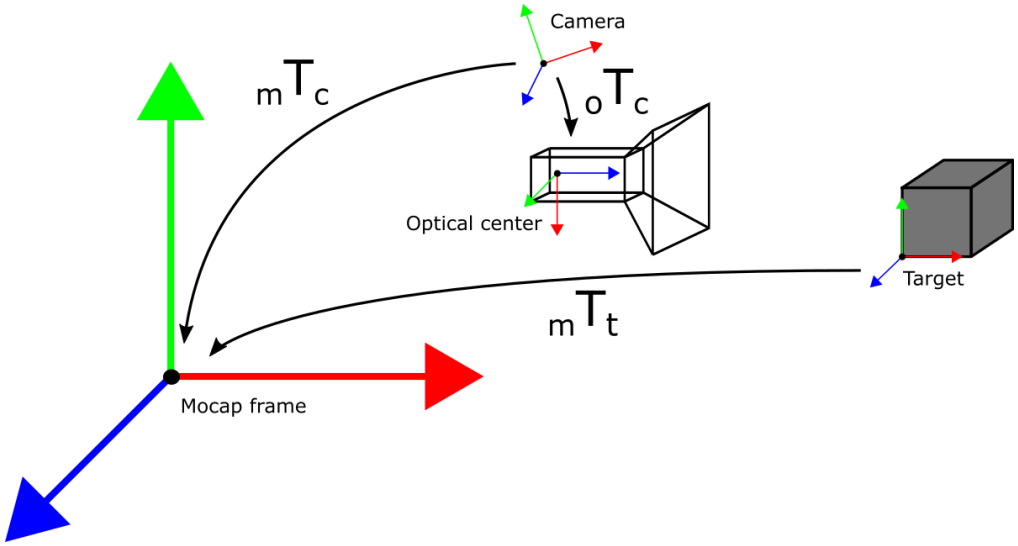


Figure 16: Coordinate frames in the MoCap system.

This kind of calibration, like camera calibration in general, is often a tedious and relatively time-consuming process. For that reason, a calibration toolkit and pipeline were created to calibrate and find transformation automatically. Additionally, a small proof-of-concept 6DOF object tracking dataset was created with RGB and depth cameras. Figure 17 shows the cameras, calibration pattern and objects that were used in the process.

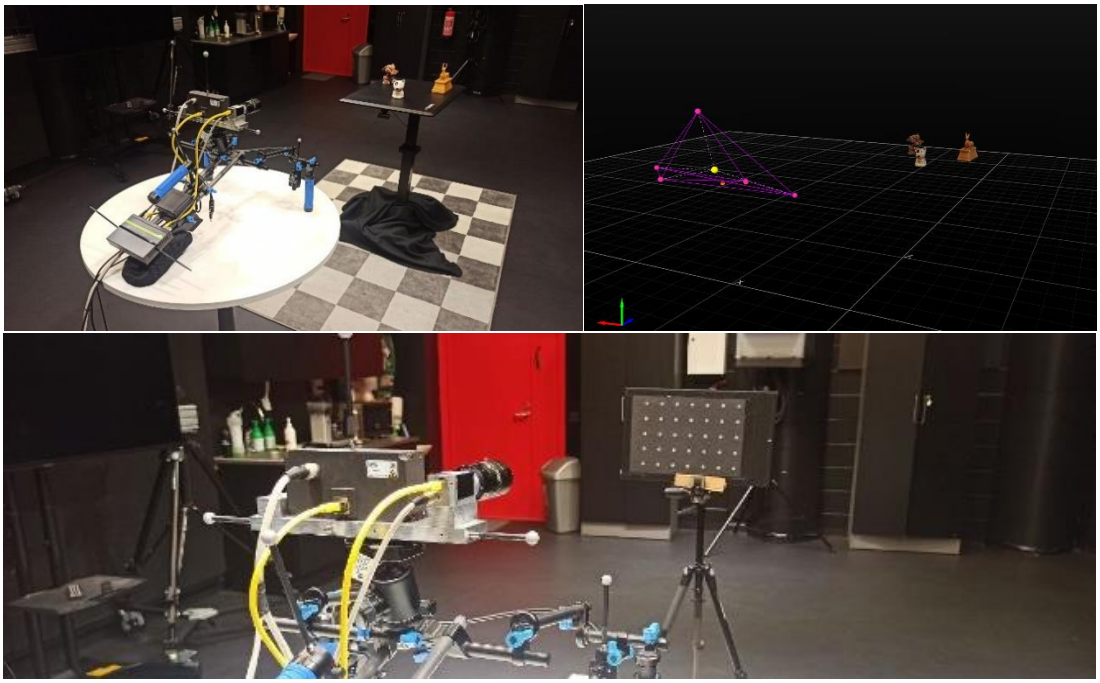


Figure 17: Camera rig and tracked objects in real world (top-left), in MoCap software (top-right) and a calibration at the bottom.

In the developed method, video take is recorded while camera is stationary. During this take, calibration pattern attached on tripod is moved around in the volume such that it is visible to the camera and MoCap system. Checker is moved to ~20 different positions and left on the ground for ~1s period to avoid any errors from motion blur and wiggling. The pose of the camera and positions of each dot on the checker are tracked in the MoCap application. This MoCap data is first utilized by detecting stationary periods in checker positions to find suitable

frames for calibration. After that the checker dots are detected from these images with the help of Python and OpenCV software. Figure 18 shows the different calibration stages.

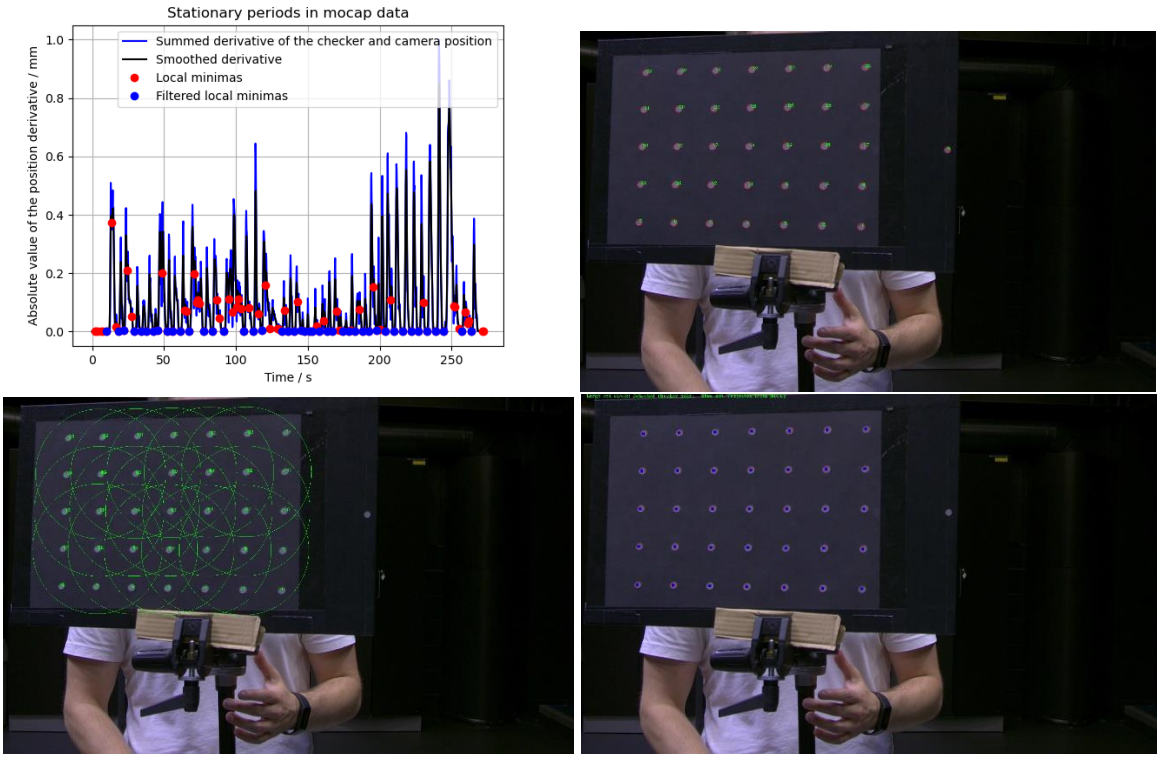


Figure 18: Stationary frame finding, checker pattern detection and verification.

Now intrinsic matrix and camera poses are estimated with OpenCV's camera calibration function. With intrinsic matrix and two sets of points from different sources, it is possible to estimate transformation ${}_{o}\mathbf{T}_{c}$ with a least-squares method. This transformation can then be used to transform points from the MoCap coordinate frame to the camera coordinate frame. In Figure 19, a 3D model of an ape is transferred to the camera's coordinate frame by using MoCap data and calibrated transformation. For visualization purposes the model was then projected onto the image plane and overlayed with blue tint. The 3D model was initially manually aligned with the corresponding object in MoCap system. This would not work with a camera but here the system is not as sensitive to small errors in offset.

Figure 19: Example frames from the dataset.

Open questions and research directions

Exact synchronization between the camera and MoCap system is needed. Here it was implemented with an electrical trigger signal, but that is not always possible. And even when it is, it takes some effort to implement for different cameras. An easier out-of-the-box method

would be useful. Another research avenue would be to create an implementation for other objects than a camera. For example, this could be done for a VR headset where object pose can be accessed from the VR software instead of camera calibration algorithm.

Finally, a robust and user-friendly way to align real objects with objects in the MoCap system would be valuable for various applications. Essentially this is same problem as calibrating transformation for a camera or VR headset, but some different approach should be chosen.

References

- https://optitrack.com/
- https://www.movella.com/

Conclusions

This white paper presented CIVIT technologies that were profiled as a part of the Collaboratec project: volumetric capture, depth and hyperspectral sensors and motion capture. The findings can help to increase awareness of the CIVIT infrastructure and potentially establish new collaboration with companies where the technology is utilized in companies' own projects. Due to the needs of the work machine and creative digital industries, the technology profiling and subsequently this white paper focused predominantly on different sensing and capture technologies. In the future, the document can be extended to also include other CIVIT technologies such as immersive 3D displays and mobile robotic platforms.

Collaboratec: Collaboration model for using immersive visual technologies in work machine and creative digital industries. Project code: A77587.
https://www.tuni.fi/en/research/collaboratec

CIVIT: Centre for Immersive Visual Technologies at Tampere University
www.civit.fi
civit@tuni.fi

European Union
European Regional Development Fund

Leverage from the EU 2014–2020

PIRKANMAA