WHOLODANCE

Whole-Body Interaction Learning for Dance Education

Call identifier: H2020-ICT-2015 - **Grant agreement no**: 688865

Topic: ICT-20-2015 - Technologies for better human learning and teaching

Deliverable 2.8

Multi-sensor integration report

Due date of delivery: August 31st, 2016

Actual submission date: September 2nd, 2016

Start of the project: 1st January 2016
Ending Date: 31st December 2018

Partner responsible for this deliverable: LYNKEUS

Version: 3.0

Dissemination Level: Public

Document Classification

Title	**Multi-sensor integration report**
Deliverable	D2.8
Reporting Period	M1-M8
Authors	Oshri Even-Zohar
Work Package	WP2
Security	Public
Nature	Report
Keyword(s)	Sensor, Motion Capture, Devices' Integration,

Document History

Name	**Remark**	**Version**	**Date**
Oshri Even-Zohar	First Version	1.0	20th August 2016
Oshri Even-Zohar	Final Version with updated data from new devices	2.0.	1st September
Stefano Di Pietro	Final Revision and Submission	3.0.	2nd September

List of Contributors

Name	**Affiliation**
Oshri Even-Zohar	MOTEK

List of reviewers

Name	**Affiliation**
Stefano Di Pietro	Lynkeus
Antonella Trezzani	Lynkeus

Table of contents

1 Purpose of the document

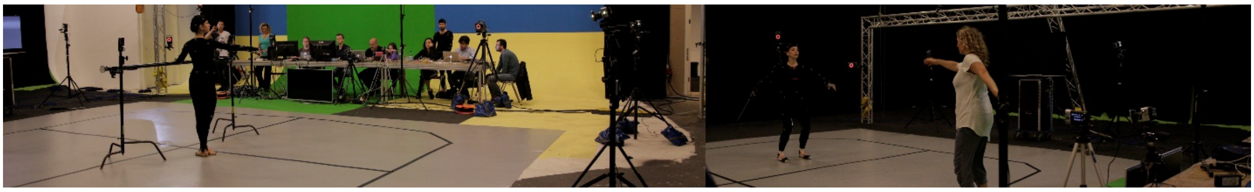
This report contains some preliminary results of integration of suggested additional registration devices. The issues encountered during the task and a set of recommendations for the optimal usage of multimodal sensors

2 The base template specification

Motion capture data base template was created using Vicon™ optical motion capture systems. Those systems provide high fidelity, high accuracy and high sampling rate. All other sensor types will be compared in this document to the Vicon™ base template specification.

2.1 Capture Volume

The Capture volume achieved was an ellipsoid of approx. 9.5 X 7 X 3.8 meters. This volume was used for all 4 dance Genres captured in Amsterdam during May and July 2016.



2.2 System Accuracy

The accuracy reached was between 0.7 and 1.5 mm. This is measured through averaging of all the RMS (Residual Margin Errors) of the motion capture systems after each system calibration (Twice daily)

2.3 Sampling speeds

For both sessions, all motion capture data was sampled at 120Hz. We have also created some tests of Classic dance motions (Pirouettes) in high speed (500Hz) to be able to view it in slow motions.

2.4 Processed data accuracy

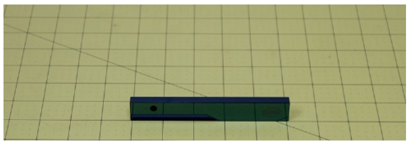
The measurements of this parameters are done separately for body-part / joint translations and rotations. The translation accuracy is 0.2-0.3 mm. Rotation accuracy is between 2 and 5 minutes of a degree. This measurement is also verified by comparisons to selected sequences video footage.

3 Suggested additional registration devices.

There are several comparable sensing devices that use different paradigms. (We only count devices that create different data formats of 3D data. 2D data registration (like Video) were not taken in consideration. Those are: Depth cameras, ToF (Time of Flight) depth cameras, Inertial sensing devices (Xsens, Isense, IMU sensors), Standalone accelerometers and ElectroMagnetic devices (Assencion, Polhemus).

3.1 Depth cameras

Among the devices that were compared in this category are: MS Kinect1, MS Kinect2, Intel Real-sense R200 and F200, Orbbec Astra, DUO mini lx and ZED stereo cam.



The original Kinect sensor software by Microsoft is still supported, but the hardware is discontinued early in 2015. The sensor works indoors to a range of about 4.5m and can track the skeletons of two people simultaneously. The official SDK supports only Microsoft platforms, but the community has implemented support for other frameworks. The sensor connects via USB 2 and requires its own power source. It runs at 30Hz only.

The second generation of the Kinect family is the Kinect2— it's physically the largest sensor, and it requires a dedicated USB 3.0 bus and its own power source. For all that, you get a wider field of view and relatively cleaner depth data at a range of .5m-4.5m, further away the data becomes very noisy. The SDK, provides full skeleton tracking of six people simultaneously, basic hand open/close gestures, and face tracking. Microsoft provides a plugin for Unity 3D. On the downside, it's tough to extend the device very far from the host computer, you can only use one sensor per computer, and only on Windows 8 and above.

Intel's RealSense devices (https://software.intel.com/en-us/intel-realsense-sdk) are meant to be integrated into to OEM products, but the developer toolkits are available for use in installation projects. The R200 is the second RealSense product to ship from Intel, and it's a tiny USB 3 device with an infrared sensing range of about .5m-3.5m. The "R" is for rear-facing, meaning its primary use case is to be integrated into the back of a tablet or laptop display. The SDK is quite robust, supporting C++, C#, JavaScript, Processing, Unity, and Cinder. The SDK supports face and expression tracking, but not hand tracking or full skeletons. The device really comes into its own when the camera in motion for augmented reality or 3D scanning applications.

The Intel Realsense F200https://software.intel.com/en-us/RealSense/F200Camera is meant to be front-facing, and excels at tracking faces, hands, objects, gestures, and speech. It's meant to be mounted to the front of a display or tablet and has a sensing range of about 0.2m-1.2m and a 60FPS VGA depth stream. The SDK is quite robust, supporting C++, C#, JavaScript, Processing, Unity, and Cinder.

Orbbec is a relatively new 3D camera device. Their first products are the Astra and Astra Pro, which are both infrared depth sensors with a 640×480 resolution at 30FPS. The SDK is supporting only the older C++ OpenNI framework. Support Unity 3D is said to be forthcoming. The SDK supports basic hand tracking which can be used for gestural interfaces, but not full skeleton tracking. The unit can sense as far as 8 meters away, which beats the range of most other sensors.

The ZED camera from StereoLabs is unique among this list as it does not use infrared light for sensing, but rather a pair of visible light sensors to produce a stereo image, which is then delivered to software as a video stream of depth data. It works well outdoors to a depth of 20 meters and provides a high-resolution depth image of up to 2208×1242 at 15FPS, or VGA at 120FPS. While the hardware is quite powerful, the provided SDK is pretty limited to simply capturing the depth stream, without any higher-level interpretation. Any tracking of objects, hands, faces, or bodies would need to be implemented by the developer.

The DUO mini lx camera is a tiny USB-powered stereo infrared camera that provides high-frame-rate depth sensing to a range of about 3m. It includes IR emitters for indoor use, but can be run in a passive mode to accept ambient infrared light — meaning it can be used outdoors in sunlight. The SDK provides a basic depth map via a C interface, but no higher-level tracking of hands, faces, or skeletons. It does however work on OS X and Linux, and even ARM-based systems.

Below is a comparison table looking at the differences in specifications.

	Orbbec Astra	RealSense R200	ZED Stereo Camera	RealSense F200	Kinect for XBox One	DUO mini lx	Kinect for XBox 360
Released	September 2015	September 2015	May 2015	January 2015	July 2014	May 2013	June 2011
Price	$150	$99	$449	$99	$100	$695	Unavailable
Tracking Method	IR	IR	Stereo RGB cameras	IR	IR	Passive IR	IR
Range	0.4m to 8m	0.5m – 3.5m	1.5m – 20m	0.2m – 1.2m	0.5m – 4.5m	0.3m – 2.4m	0.4m – 4.5m
RGB Image	1280×960, 10 FPS	1920×1080, 30 FPS	configurable between 1280×480, 120 FPS and 4416×1242, 15 FPS	1920×1080, 30 FPS	1920×1080, 30 FPS	configurable between 320×120, 360 FPS and 752×480, 56 FPS	640×480, 30 FPS
Depth Image	640×480, 16 bit, 30 FPS	640×480, 60 FPS	configurable between 640×480, 120 FPS and 2208×1242, 15 FPS	640×480, 60 FPS	512×424, 30 FPS	configurable between 320×120, 360 FPS and 752×480, 56 FPS	320×240, 30 FPS
Connectivity	USB 2.0	USB 3.0	USB 3.0	USB 3.0	USB 3.0	USB 2.0	USB 2.0
Physical Dimensions	160×30×40 mm	130×20×7 mm	175×30×33 mm	150×30×58 mm	250×66×67 mm	52×25×11 mm	280×64×38 mm
Works outdoors?	✗	✗	✓	✗	✗	✓	✗
Skeleton tracking?	✓ (only hand positions)	✗	✗	✗	✓ (six skeletons)	✗	✓ (two skeletons)
Facial tracking?	✗	✓	✗	✓	✓	✓	✓
3D scanning?	✗	✓	✓	✓	✓	✓	✓

Simultaneous apps?	✗	✓	✗	✓	✓	✗	✗
Gesture Training?	✗	✗	✗	✗	✓ (Visual Gesture Builder)	✗	✗ (only via third-party tools)
Gesture Detection?	✗	✗	✗	✓	✓ (hand open, closed, lasso)	✗	✓ (hand grip, release, press, scroll)
Toolkits	OpenNI	Java, JavaScript, Processing, Unity3D, Cinder		Java, JavaScript, Processing, Unity3D, Cinder	WPF, Cinder, OpenFrameworks, JavaScript, vvvv, Processing, Unity3D, more	Dense3D, OpenCV, Qt5	WPF, Cinder, OpenFrameworks, JavaScript, vvvv, Processing, Unity3D, more
Languages	C++/OpenGL	C++, C#, Java, JavaScript	C++/OpenGL	C++, C#, Java, JavaScript	C#, C++, JavaScript, Java	C++, C#	C#, C++, JavaScript, Java
Project Examples	HandViewer, Depth Data Viewer, RGB Data Viewer	Face tracking examples C++. Only one Unity3D sample.	Background subtraction, right image disparity, depth map	Many examples of face tracking, gesture tracking, speech detection on a variety of different platforms and frameworks	Many examples of skeleton tracking, on a variety of different platforms and frameworks	Very few samples in each of the supported languages, mostly to get raw image and depth data	Many examples of skeleton tracking, face tracking, and speech detection on a variety of different

3.2 Inertial sensing devices

There are many manufacturers of standalone IMU sensors. A Typical IMU sensor will contain accelerometers, Gyroscopes and magnetometers. For a wide list of IMU manufacturers see:

http://damien.douxchamps.net/research/imu/

We have only concentrated on IMU systems that are capable of tracking full body. I.e. Multiple synchronized IMU based systems. There are currently only 4 systems that answer such specification. Those are: Xsens, Intersense, Synertial and 3Dsuit. There are several low cost new systems (like Noitom) on the market and there are several new systems in development in different parts of the world.

Below is a rough comparison table between inertial tech and optical tech.

	Optical	**Inertial**
Capture Area	Large	Large
Occlusions	Yes	No
Calibration	Easy	Complex
Data Cleaning	Auto	Manual
Sensor placement	Easy	Easy
Motion Accuracy	Very high	High
Latency	None	Slight

Lag	None	Slight
Synchronicity	Yes	yes
Sampling rate	>120Hz	>120Hz
Floor contact	Yes	yes
Drift accumulation	No	Yes

3.3 Electromagnetic systems

The comparison between optical and Electromagnetic systems counts the following parameters:

- Benefits of Optical system
    - Very clean and detailed data
    - Cable-less setup; allows performer more freedom of movement
    - Large data capture area: can track multiple subjects and more complex performances
    - High sampling rates
    - High data capture volume
- Drawbacks of Optical system
    - Extremely high costs ($150K-$250K)
    - Prone to interference from light, reflections, or physical objects
    - Marker occlusion can interfere with data collection (can be compensated for with software which estimates the position of a missing dot)
    - Originally not real-time due to post-processing of data; advances have made real-time optical tracking possible, though there is greater chance for latency when compared to mechanical and electromagnetic systems

In the electromagnetic system, performers don an array of magnetic receivers. These calculate position and orientation via the relative magnetic flux of three orthogonal coils on the receivers that the performer wears and a static magnetic transmitter. The relative intensity of the voltage or current of the three coils allows these systems to calculate both range and orientation by meticulously mapping the tracking volume.

One of the first uses of electromagnetic motion capture was for the military, to track head movements of pilots. Often this method is layered with animation from other input devices.

- Benefits of magnetic systems
  - Real-time tracking
  - Can capture data with only a fraction of the markers compared to optical
  - More absolute data: positions/rotations are measured absolutely, orientation in space can be determined
  - Not occluded by nonmetallic objects
  - Relatively cheaper than optical
- Drawbacks of magnetic systems
  - Data is noisy (less clean) compared to optical
  - Prone to magnetic and electrical interference (rebar, wiring, lights, cables, etc.)
  - Restricted data capture area: performers wear cables connecting them to a computer, which limits their freedom of motion
  - Significantly lower sampling rate and data capture volume compared to optical