INTERNATIONAL STANDARD

ISO/IEC 18520

First edition
2019-01

Information technology — Computer graphics, image processing and environmental data representation — Benchmarking of vision-based spatial registration and tracking methods for mixed and augmented reality (MAR)

Technologies de l'information — Infographie, traitement d'images et représentation des données environnementales — Étalonnage des méthodes d'enregistrement géométriques et de suivi basées sur la vision pour le MAR

Reference number
ISO/IEC 18520:2019(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see http://patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 24, *Computer graphics, image processing and environmental data representation*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

In the development of mixed and augmented reality (MAR) applications, one of the most important technologies involves spatial registration and spatial tracking methods, especially vision-based methods. The research and development on registration and tracking based on computer vision technologies is flourishing and many new algorithms have been proposed every year.

Therefore, this document aims at fostering objective evaluation and comparison of diverse registration and tracking methods, in order to facilitate fairer competition among small and major companies/institutes involved in MAR technologies, applications and services.

Moreover, this document can be the baseline to standardize spatial registration and tracking methods which not only utilize a video camera but combine a video camera with other sensors such as another video camera, a depth camera, inertial sensors and infrastructure-based positioning technologies, and which utilize technologies such as IoT (Internet of Things) and GNSS (Global Navigation Satellite System).

The target audience of this document includes stakeholders of benchmarking activities. The following are examples of how this document can be used directly or indirectly:

— by a benchmarking service provider, a benchmark provider or a benchmarking competition organizer who wishes to align their benchmarking activities including self-benchmarking and open/closed competitions to be consistent with this document;

— by a technology developer/supplier who wishes to estimate and evaluate the performance of a vision-based spatial registration and tracking (vSRT) method for MAR appropriately with a benchmarking service provider, a benchmark provider or a benchmarking competition organizer who aligns their benchmarking activities to be consistent with this document; or

— by a technology user who wishes to obtain benchmarking results based on a benchmarking activity, which is consistent with this document, or to compare the existing vSRT methods for MAR in terms of their performance.

Information technology — Computer graphics, image processing and environmental data representation — Benchmarking of vision-based spatial registration and tracking methods for mixed and augmented reality (MAR)

1 Scope

This document identifies the reference framework for the benchmarking of vision-based spatial registration and tracking (vSRT) methods for mixed and augmented reality (MAR).

The framework provides typical benchmarking processes, benchmark indicators and trial set elements that are necessary to successfully identify, define, design, select and apply benchmarking of vSRT methods for MAR. It also provides definitions for terms on benchmarking of vSRT methods for MAR.

In addition, this document provides a conformance checklist as a tool to clarify how each benchmarking activity conforms to this document in a compact form by declaring which benchmarking processes and benchmark indicators are included and what types of trial sets are used in each benchmarking activity.

2 Normative references

There are no normative references in this document.

3 Terms, definitions, acronyms and abbreviated terms

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1 Terms and definitions

3.1.1
benchmark
reference point against which comparisons can be made

Note 1 to entry: In the context of this document, the performance of vSRT methods for MAR is the object of comparison.

Note 2 to entry: See ISO/IEC 29155-1.

3.1.2
benchmark indicator
indicator that qualitatively shows a particular aspect of a *benchmark* (3.1.1) with appropriate metrics

3.1.3
benchmarking
activity of comparing objects of interest to each other or against *benchmarks* (3.1.1) to evaluate relevant characteristics

Note 1 to entry: In the context of this document, the object of interest is the performance of vSRT methods for MAR, and the characteristics are particular aspects of the performance such as reliability, temporal characteristic, etc.

Note 2 to entry: See ISO/IEC 29155-1.

3.1.4
benchmarking instrument
tool, method or guide used to support every activity within the *benchmarking* (3.1.3) framework

3.1.5
benchmarking method
particular procedure for conducting *benchmarking* (3.1.3) and obtaining *benchmarking results* (3.1.7)

3.1.6
benchmarking repository
repository which is designated for retaining information necessary for *benchmarking* (3.1.3) such as *datasets* (3.1.9) for benchmarking and *benchmarking results* (3.1.7)

Note 1 to entry: Some benchmarking repositories might contain all information, and some might contain a subset.

Note 2 to entry: See ISO/IEC 29155-1.

3.1.7
benchmarking results
benchmarks (3.1.1) as primary results, and intermediate results and reports on *benchmarking* (3.1.3) including findings, issues and lessons learned as secondary results

Note 1 to entry: In the context of this document, a typical intermediate result is the output of vSRT methods such as the estimated position of a camera the accuracy of which is evaluated with a benchmark indicator.

3.1.8
competition organizer
person or organization that hosts an *on-site* (3.1.14) or *off-site* (3.1.13) benchmarking competition

3.1.9
dataset
collection of data that contains target images and the *ground truth* (3.1.11) regarding the target images for *benchmarking* (3.1.3)

3.1.10
extrinsic camera parameters
parameters that define the position and orientation of a camera reference frame with respect to a known world reference frame such as the translation vector and the rotation matrix

3.1.11
ground truth
collection of measurements that is much more accurate as a whole than measurements by technologies which are the targets of *benchmarking* (3.1.3)

3.1.12
intrinsic camera parameters
parameters that define the relationship between the pixel coordinates of an image point and the corresponding coordinates in the camera reference frame

Note 1 to entry: Intrinsic camera parameters contain the focal length, the scale factors, the skew, the principal point, the lens distortion, etc.

3.1.13
off-site benchmarking
benchmarking (3.1.3) that is conducted with target images in datasets which were prepared beforehand

3.1.14
on-site benchmarking
benchmarking (3.1.3) that is conducted by executing vSRT programs while capturing target images on the spot

3.1.15
physical object
object, which exists in the real world, used as a target of *spatial registration* (3.1.16), *spatial tracking* (3.1.17) and/or augmentation

Note 1 to entry: Especially for off-site benchmarking with physical objects, the physical object instances shall be easily available or deliverable objects such as paper crafts and toy bricks, or they shall be made accessible by providing information on how to find them for capturing the images. The physical object instances in off-site benchmarking are utilized to gather and acquire information necessary for vSRT methods such as visual features and 3D models of the objects.

3.1.16
spatial registration
establishment of the spatial relationship or mapping between two models, typically between virtual objects and target *physical objects* (3.1.15)

[SOURCE: ISO/IEC 18039:2019, 3.1.20]

3.1.17
spatial tracking
update of the spatial relationship or mapping between two models, typically between virtual objects and target *physical objects* (3.1.15) over time

3.1.18
trial set
combination of a dataset and a collection of *physical object* (3.1.15) instances for *off-site* (3.1.13) and *on-site* (3.1.14) benchmarking of vSRT methods for MAR

3.1.19
vision-based spatial registration and tracking
spatial registration (3.1.16) and *spatial tracking* (3.1.17) based on image processing and computer vision technologies

Note 1 to entry: The term “spatial registration and tracking” is also called “geometric registration and tracking” in Annex A and in some of the bibliography references.

3.2 Acronyms and abbreviated terms

3DEVO	3D error of a virtual object
MAR	Mixed and augmented reality[1]
PEVO	Projection error of a virtual object
SLAM	Simultaneous localization and mapping[1]
vSRT	Vision-based spatial registration and tracking

4 Overview of the framework

This clause outlines the reference framework of benchmarking of vSRT methods for MAR, the details of which are described in Clauses 5, 6, and 7. As shown in Figure 1, the reference framework is composed of the following three core components.

— **Benchmarking processes**, which include how to produce benchmarking outcomes such as benchmarking results, benchmark surveys and benchmarking instruments with benchmark indicators and trial sets and how to share benchmarking outcomes.

— **Benchmark indicators**, which quantify the performance of vSRT methods in MAR by taking into account not only characteristics of vSRT methods in MAR such as reliability and temporal characteristics, but also fair comparisons.

— **Trial set elements**, which are composed of datasets and physical object instances to provide each benchmarking attempt with the same conditions.

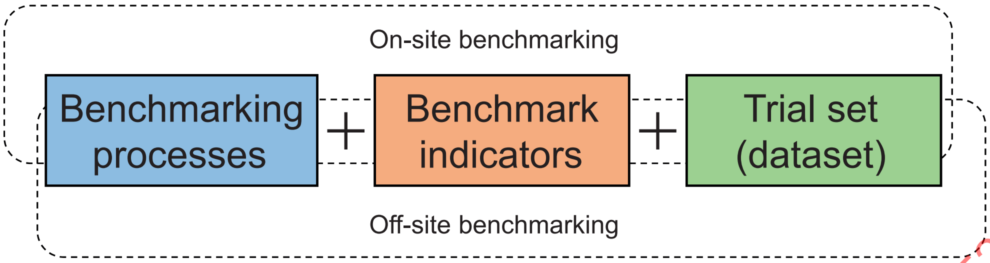


Figure 1 — Core components of on- and off-site benchmarking framework

The above three components are identified and defined in accordance with grass-roots activities for standardizing benchmarking schemes and for conducting on-site or off-site comparison of vSRT methods and MAR systems which are often held as contests and are introduced in Annex A.

On-site benchmarking methods are used to conduct benchmarking on the spot while capturing images of physical objects with working MAR systems. Compared with off-site benchmarking methods mentioned afterwards, it is inevitable that human factors affect on-site benchmarking results due to time and cost limitations, and constraints in preparation and operation. Therefore, it is highly recommended to simplify the implementation of benchmarking frameworks for on-site benchmarking.

By contrast, off-site benchmarking methods are used to conduct benchmarking with target images in datasets prepared beforehand. Compared with on-site benchmarking methods, the stakeholders have more time for preparing and conducting benchmarking. However, additional effort on the implementation of benchmarking frameworks is needed for alleviating issues related to fine tuning and cheating.

The on-site or off-site competition is one of the most concrete cases of on-site or off-site benchmarking methods, respectively. A.6, A.7 and A.8 introduce several case examples of on- and off-site competitions.

Typical processes of on- and off-site benchmarking are extracted from the grass-roots activities as in Annex A, and they are schematically described in Clause 5 by referring to the ISO/IEC 29155 series, especially ISO/IEC 29155-1. The conceptual relationship among the ISO/IEC 29155 series, ISO/IEC 18039 and this document is graphically indicated in Annex C. Each benchmark indicator and trial set element is also extracted from outcomes and discussions in the grass-roots activities as in Annex A. Clause 6 describes three major types of benchmark indicators which correspond to reliability, temporality and variety indicators, and Clause 7 describes reference elements in a trial set which contains dataset elements and physical object instances.

5 Benchmarking processes

5.1 Overview

This clause outlines benchmarking processes and related components necessary to produce and share benchmarking outcomes. Figure 2 illustrates the basic benchmarking process flow.

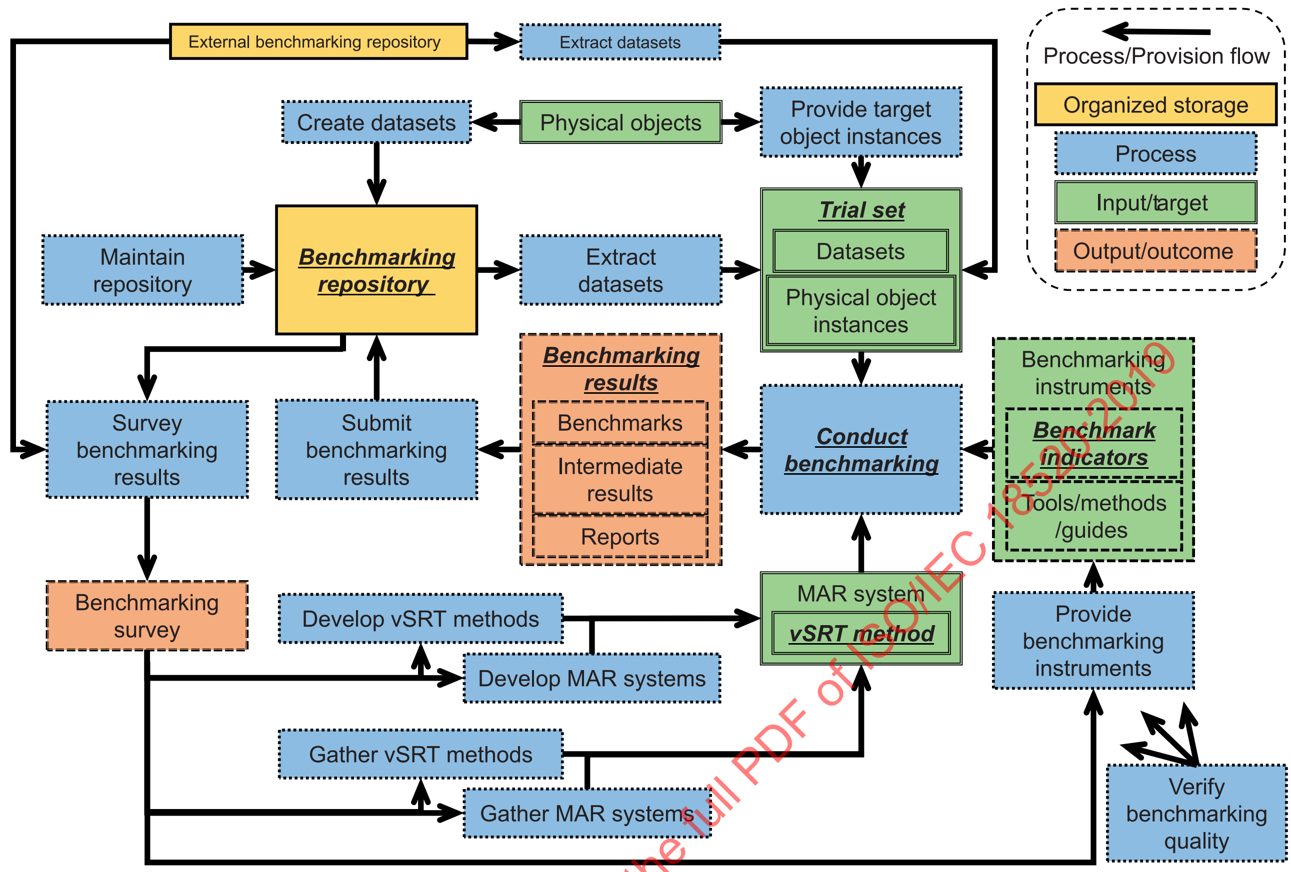


Figure 2 — Basic benchmarking process flow

5.2 Process and process flow

Although the details of the process flow in Figure 2 can differ in each specific benchmarking, it generally consists of process, target, input, output/outcome and organized storage, described in detail as follows:

— **Process**, which consists of one or more micro processes:

— to develop or gather vSRT methods and MAR systems,

— to prepare or conduct benchmarking,

— to provide or maintain benchmarking instruments and repositories,

— to share benchmarking results,

— to manage or verify benchmarking quality.

— **Target**, which is a vSRT method or MAR system used as a benchmarking target.

— **Input**, which includes:

— trial sets,

— physical objects,

— benchmarking instruments such as benchmark indicators, tools, methods and guides.

NOTE Trial sets and benchmarking instruments are also regarded as important outcomes of benchmarking activities.

— **Output/outcome**, which includes:

  — benchmarking results such as benchmarks, intermediate results and reports,

  — benchmarking surveys.

— **Organized storage**, which is a benchmarking repository or other external repository.

5.3 Stakeholders

Various stakeholders are involved in processes on benchmarking of vSRT methods for MAR. Figure 3 illustrates a typical example of the correspondence between stakeholders and their roles in benchmarking processes. Based on roles, stakeholders can be logically classified into the following groups[2]:

— **Benchmark provider**, who creates and gathers datasets, maintains benchmarking repositories and provides benchmark surveys;

— **Benchmarking service provider**, who develops and provides benchmarking instruments, prepares trial sets, conducts benchmarking at the request of technology users and submits benchmarking results to a benchmarking repository;

— **Quality verifier**, who verifies benchmarking quality;

— **Technology developer**, who develops vSRT methods or MAR systems;

— **Technology supplier**, who supplies vSRT methods or MAR systems that technology developers have developed;

— **Technology user**, who chooses and utilizes vSRT methods or MAR systems based on the outcomes of benchmarking.

Targets of benchmarking are vSRT methods or MAR systems developed by technology developers or gathered by technology suppliers. To conduct benchmarking, benchmarking service providers shall prepare benchmarking instruments and trial sets. Datasets in the trial sets are extracted from a benchmarking repository. For on-site benchmarking, physical objects including rooms and spaces shall also be prepared. Benchmarking of vSRT methods or MAR systems is conducted with those inputs, and the results of benchmarking are submitted in a benchmarking repository by benchmarking service providers.

To choose appropriate vSRT methods or MAR systems, technology users refer to benchmark surveys or benchmarking results. The quality of the benchmark surveys or benchmarking results shall be ensured by verifying the quality of processes, inputs, outputs and organized storages in benchmarking activities. This is the main role of quality verifiers.

Various role-sharing schemes can be used in practice. Any person or organization can fulfil one or more roles. For example, benchmark providers can also have a role as benchmarking service providers. By contrast, one role can be fulfilled by several persons or organizations. For example, the competition organizer together with the contestants often fulfil the role of benchmarking service provider in conducting benchmarking. Many academic researchers do not maintain benchmarking repositories for a long term, but they often publish benchmarking surveys by conducting benchmarking for comparing several vSRT methods. In this case, they partially fulfil the benchmarking service provider's role and benchmark provider's role. In other cases, technology developers and suppliers often fulfil the partial roles of a benchmarking service provider in self-benchmarking or of a contestant of an on-site or off-site benchmarking competition.

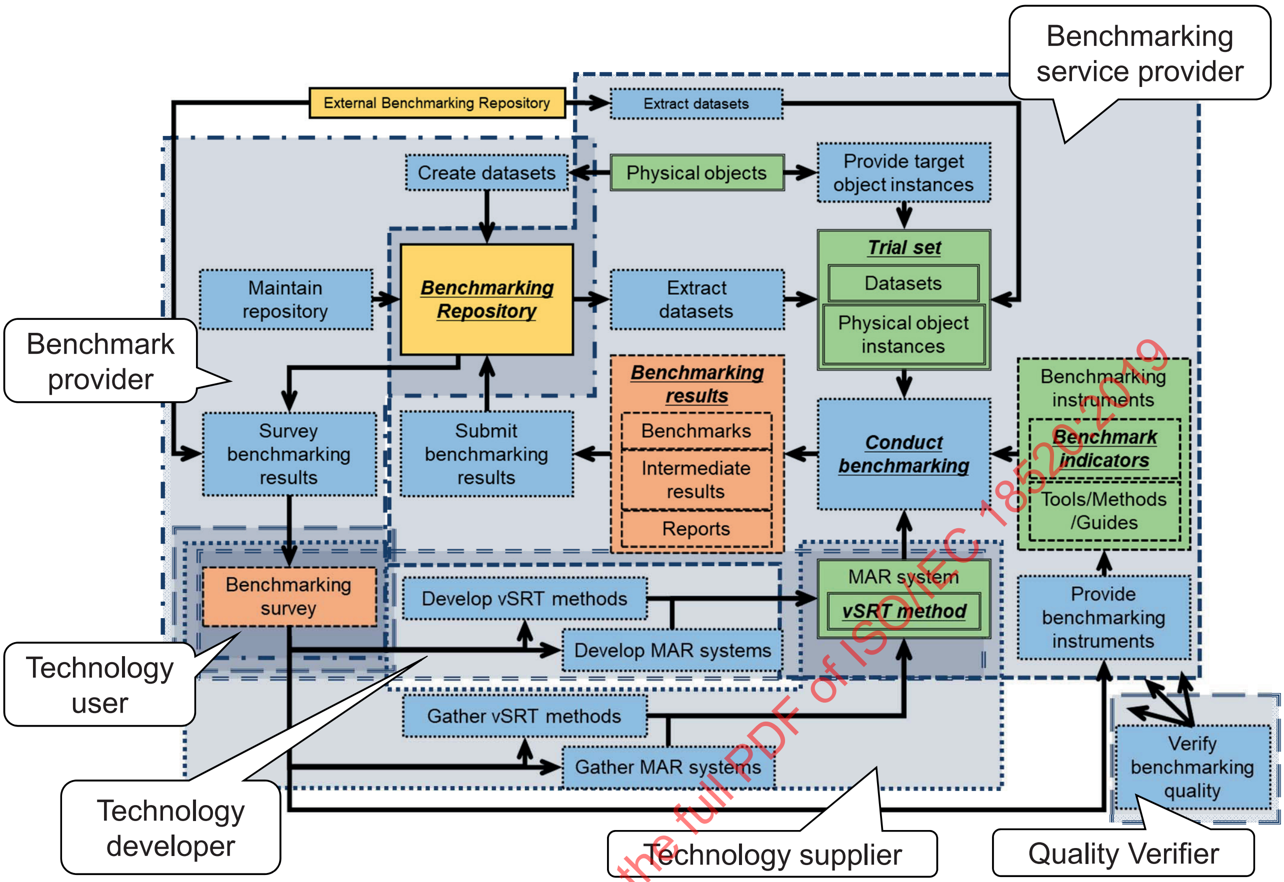


Figure 3 — Example of the correspondence between stakeholders and their roles in benchmarking processes

6 Benchmark indicators

6.1 Overview

This clause outlines three major types of benchmark indicators (reliability, temporality and variety), which should be utilized for fair comparison of vSRT methods in MAR. Table 1 shows representative benchmark indicators for both on- and off-site benchmarking.

Table 1 — Benchmark indicators for off-site and on-site benchmarking

	Off-site	On-site
Reliability	— 3DEVO — PEVO — Re-projection error of image features — Position and posture errors of a camera	— 3DEVO — PEVO — Re-projection error of image features — Position and posture errors of a camera — Completeness of a trial
Temporality	— Throughput — Latency	— Throughput — Latency — Time for trial completion
Variety	— Number of datasets — Variety on properties of datasets	— Number of trials — Variety on properties of trials

6.2 Reliability indicators

This subclause presents reliability indicators. The following four indicators are for both off-site and on-site benchmarking.

— **3D error of a virtual object (3DEVO)**, which is the difference between the estimated position of a virtual object and the ground truth. 3DEVO is one of the most direct and intuitive indicators for vSRT methods for MAR, as one of the principal functions of MAR systems is to align virtual objects in 3D space based on the results obtained by the target vSRT method.

— **Projection error of a virtual object (PEVO)**, which is also one of the most direct and intuitive indicators for vSRT methods for MAR, as one of the most important functions of MAR systems is to render virtual objects based on the results obtained by the target vSRT method[3]. Assuming the simplest case in which a virtual point is projected as a virtual object to an estimated image plane, the distance between the projected and correct points is calculated as a PEVO value. The PEVO value can be measured in degrees or in pixels. Angular distance measure can provide a uniform measure in a screen space, whereas pixel number varies depending on positions in a screen space.

— **Re-projection error of an image feature**, which is the distance between a detected image feature in an image plane and the re-projection to the image plane with the 3D coordinates of the image feature that are recovered based on the target vSRT method. Assuming the simplest case in which the image feature is a feature point, the re-projection error can be the distance between the detected feature point and the re-projected point, and can be measured in degrees or in pixels as with PEVO.

— **Position and posture errors of a camera**, which is the difference between the estimated position and posture of a camera and the ground truth.

In addition to the aforementioned four reliability indicators, completeness of a trial should be employed especially for on-site benchmarking. This is because, in many on-site competitions, many MAR systems cannot help but stop performing spatial registration and tracking in the middle of the trial. Completeness of a trial involves evaluating the extent of a trial completion. It is regarded as the robustness of the target vSRT method.

6.3 Temporality indicators

This subclause presents temporality indicators which are necessary to discuss real-time issues in MAR.

The following two indicators as shown in Figure 4 are generally suitable for off-site benchmarking.

— **Throughput**, which is the rate at which a target image is processed through a target vSRT method or target MAR system during a specific period. It is often called frame rate.

— **Latency**, which is the time delay produced by a target vSRT method or target MAR system.

  — For MAR-system benchmarking, the latency might be the length of time from when starting to capture a target image with the system to when rendering a virtual object based on the estimated position and posture of a camera with which the target image was captured.

  — For vSRT-method benchmarking, the latency might be the length of time from when starting to load a target image into the target method for input to when returning the estimated position and posture of a camera with which the target image was captured.

For on-site benchmarking, the time for trial completion can also be used as a temporality indicator because it is easy to measure. It is the length of time from starting a trial to finishing it.

The time for trial completion is often used in on-site competitions and can represent the overall performance of a target MAR system. However, it generally includes other aspects of the MAR system such as image capturing, virtual object rendering and human factors regarding the operator being considered.

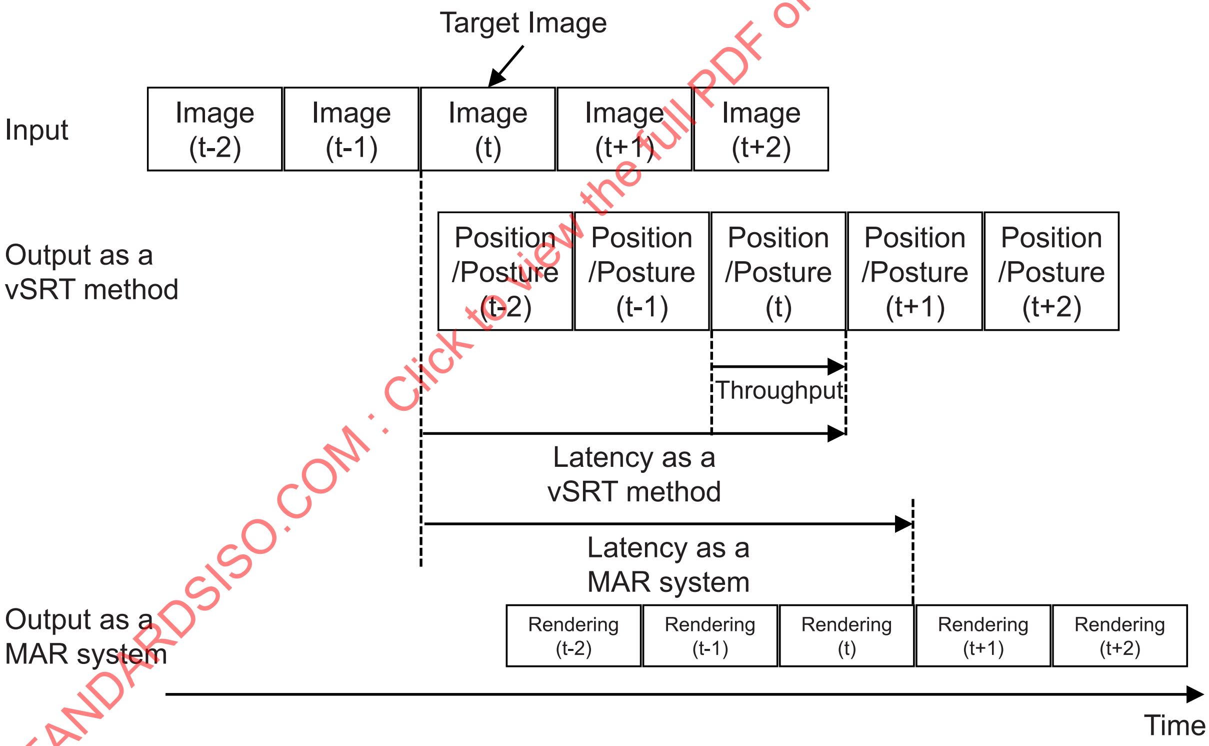


Figure 4 — Temporality indicators — Latency and throughput

6.4 Variety indicators

This subclause presents two variety indicators required to prevent fine tuning and cheating with some specific datasets for benchmarking.

The following two indicators are for off-site benchmarking.

— **Number of datasets**, which is the number of datasets used to obtain a benchmark indicator.

— **Variety on properties of datasets**, which is the variety on properties of datasets used to obtain a benchmark indicator. Typical examples of dataset properties are camera motion types, camera configurations, image quality and lighting conditions.

In general, performing fine tuning or cheating with many datasets is difficult. However, this difficulty diminishes if the properties of the datasets used for benchmarking are homogeneous.

The following two indicators are for on-site benchmarking.

— **Number of trials**, which is the number of trials attempted during on-site benchmarking.

— **Variety on properties of trials**, which is the variety on the properties of trials for on-site benchmarking.

7 Trial set for benchmarking

7.1 Overview

This clause identifies the reference elements in a trial set used for benchmarking. Table 2 shows representative elements in a trial set for on- and off-site benchmarking.

Table 2 — Trial set for benchmarking

<table>
<tr><th colspan="2"></th><th>Off-site</th><th>On-site</th></tr>
<tr><td rowspan="2">Dataset</td><td>Contents</td><td>— Image sequences
— Intrinsic/extrinsic camera parameters
— Challenge points
— Optional contents
  — 3D models for the target objects and for virtual objects
  — Image feature correspondences
  — Depth image sequences
  — Self-contained sensor data, etc.</td><td>— Challenge points
— 3D models for the target objects and for virtual objects</td></tr>
<tr><td>Metadata</td><td>— Scenario
— Camera motion type
— Camera configuration
— Image quality</td><td>— Scenario</td></tr>
<tr><td rowspan="2">Physical object instances</td><td>Contents</td><td colspan="2">— Physical objects</td></tr>
<tr><td>Metadata</td><td colspan="2">— Information on how to find the physical objects</td></tr>
</table>

7.2 Dataset for on- and off-site benchmarking

This subclause presents representative elements such as contents and metadata in datasets for off-site benchmarking. As shown in Table 2, elements in datasets for on-site benchmarking are a subset of them.

— **Contents**, which include:

  — **Image sequences**, which are target image sequences.

- **Intrinsic/extrinsic camera parameters**, which are the ground truth of intrinsic and extrinsic parameters for one or more video cameras used to capture the image sequences.
- **Challenge points**, which are points with the ground truth position for rating each MAR system based on a 3DEVO or PEVO measurement. The position of each challenge point is estimated and visualized by the target MAR system for rating purposes.
- **Optional contents**, which can be included in datasets according to requirements of each specific benchmarking. The following elements are typical examples of optional contents:
  - **3D models for both target and virtual objects**, which are 3D model data for the target objects in image sequences and for virtual objects overlaid in benchmarking.
  - **Image feature correspondences**, which is the ground truth of image feature correspondences through the image sequences.
  - **Depth image sequences**, which are image sequences where each pixel value is measured with depth sensors using, for example, an active stereo method.
  - **Self-contained sensor data**, which are sensor data measured by self-contained sensors such as an accelerometer, gyro sensor, magnetometer or barometer.

- **Metadata**, which include:
  - **Scenario**, which is the description of an application of MAR such as indoor/outdoor navigation, tabletop MAR interface or industrial application.
  - **Camera motion type**, which is the description of camera motion such as translation only, rotation only, walking motion, handheld motion or vehicle motion.
  - **Camera configuration**, which is the description of the camera configuration such as white balance, shutter type (global or rolling) and shutter speed.
  - **Image quality**, which is the description of the target-image quality such as image resolution, defocusing and motion blur.

The image format can be JPEG, PNG or Raw. A lossless format is more appropriate. The ground truth data format can be XML (eXtensible Markup Language), X3D (eXtensible 3D), JSON (JavaScript Object Notation), etc. The 3D model format can be X3D, Collada, etc. The metadata can be used as a source of variety indicators and can also correspond to the properties of each dataset for ease of dataset retrieval. The metadata format can be XML, JSON, etc.

7.3 Physical object instances for on- and off-site benchmarking

This subclause presents representative elements in physical object instances for on- and off-site benchmarking.

In off-site benchmarking, physical objects shall be easily available or deliverable so that contestants can capture the images because such physical objects are necessary for technology developers to develop and adjust their vSRT method or MAR system before conducting off-site benchmarking with target image sequences in which the physical objects are observed. In actual benchmarking activities, paper models, toy bricks, cars, etc. can be employed as easily available or deliverable physical objects, as introduced in Annex A.

In on-site benchmarking, physical objects are also used for benchmarking preparation and actual benchmarking. For developing and adjusting vSRT methods or MAR systems, the physical objects should be easily available or deliverable as in off-site benchmarking. However, when conducting the actual benchmarking, physical objects may be observable only during trials, as in Annex A.

Metadata shall be provided with physical objects, and it shall contain information on how to find or obtain the physical objects.

8 Conformance

Table 3 shows an example of a conformance checklist. This checklist or customized ones shall be used to clarify how each benchmarking activity conforms to this document in a compact form. This checklist is useful to summarize and declare which benchmarking processes and benchmark indicators are included, and what types of trial sets are used in each benchmarking activity. Annex B shows examples of how to use the conformance checklist shown in Table 3.

Table 3 — Conformance checklist

			Check	Item	Remarks
Process flow	Process		[ ]	Develop vSRT methods and/or MAR systems:	
			[ ]	Gather vSRT methods and/or MAR systems:	
			[ ]	Prepare and conduct benchmarking:	
			[ ]	Provide and maintain benchmarking instruments:	
			[ ]	Provide and maintain benchmarking repositories:	
			[ ]	Share benchmarking results:	
			Check		
	Target/ input/ output/ organized storage		[ ]	vSRT method:	
			[ ]	MAR system:	
			[ ]	Trial sets and physical objects:	
			[ ]	Benchmarking instruments:	
			[ ]	Benchmarking results:	
			[ ]	Benchmarking surveys:	
			[ ]	Benchmarking repository:	
			[ ]	External repositories:	
			Check		
Indicator	Reliability		[ ]	3DEVO:	
			[ ]	PEVO:	
			[ ]	Re-projection error of image features:	
			[ ]	Position and posture errors of a camera:	
			[ ]	Completeness of a trial:	
	Temporality		[ ]	Throughput:	
			[ ]	Latency:	
			[ ]	Time for trial completion:	
	Variety		[ ]	Number of datasets/trials:	
			[ ]	Variety on properties of datasets/trials:	
			Check		
Trial set	Dataset	Contents	[ ]	Image sequences:	
			[ ]	Intrinsic/extrinsic camera parameters:	
			[ ]	Challenge points:	
			[ ]	Optional contents:	
		Metadata	[ ]	Scenario:	
			[ ]	Camera motion type:	
			[ ]	Camera configuration:	
			[ ]	Image quality:	
	Physical object instances	Contents	[ ]	Physical objects:	
		Metadata	[ ]	How to find the physical objects:	

For on- and off-site benchmarking, in general cases such as self-benchmarking, repeatability shall be guaranteed as much as possible by the stakeholders such as benchmarking service providers and benchmark providers. For this purpose, they shall declare the specified benchmarking framework including processes, targets, inputs, outputs, organized storages, the specified formulas of benchmark indicators and the specified format and contents of trial sets to other stakeholders. The declaration with the checklist shall be based on the reference framework in Clauses 5, 6, and 7. In addition, the stakeholders shall consistently provide trial sets and benchmarking surveys while continuously maintaining repositories as in Figure 2.

For on- and off-site benchmarking in competitions, it is more difficult to guarantee repeatability completely owing to human factors. For on-site benchmarking, condition changes over time are also inevitable. However, the competition organizers at least shall make every effort to declare the specified processes and benchmark indicators, and the format of trial sets, to other stakeholders such as the contestants as early as possible. The declaration with the checklist shall be also based on the reference framework in Clauses 5, 6, and 7.

Annex A
(informative)

Benchmarking activities[1)]

A.1 Open datasets by an academic community

In the TrakMark working group[4][5][6], various activities involving evaluations of tracking methods for MAR have been held. One of the main activities is to generate and provide datasets for benchmarking of vSRT methods. The datasets are composed of camera images and ground truth data that include intrinsic and extrinsic camera parameters of each image. Three packages were prepared as a trial, referred to as the image sequence set No.1 (Figure A.1). This set consists of indoor, outdoor and CG-based sequence packages. For example, the upper images are samples of the sequence in the film studio package. The special feature of this package is that these sequences were recorded with an HD camera and reference data were measured by physical sensors such as a rotary encoder. The lower images are samples of the sequence in the conference venue package, which includes the ground truth of the camera pose. The venue of ISMAR 2009 was recreated as a 3D CG model, and computer-generated images and the related ground truth were made from it.

Figure A.1 — Examples in TrakMark datasets No.1

Based on the experience of making the first set, five more packages as the second image sequence set were prepared. In making this set, a variation of the image sequences was focused on. For example, NAIST campus package 02 (Figure A.2) includes 18 sequences. Although these sequences were taken in the same location, there are a variety of properties as in Table A.1.

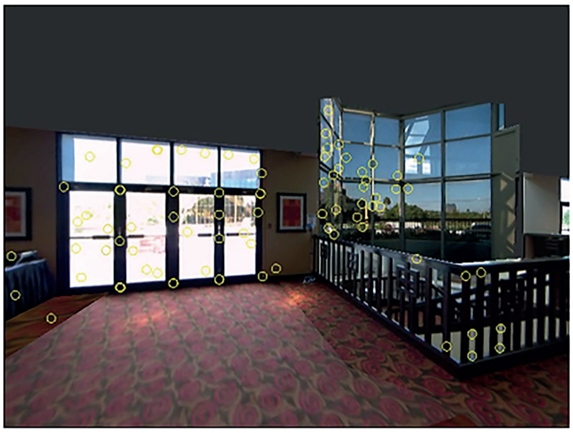
1) Any names of specific applications, file formats, etc. are given for the convenience of users of this document and do not constitute an endorsement by ISO/IEC.

Figure A.2 — Sample images of NAIST campus package 02

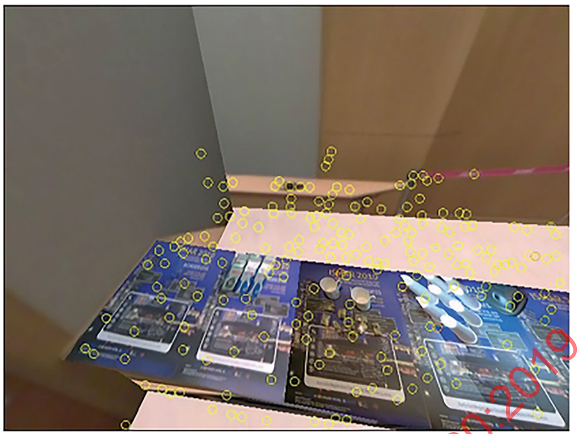
Table A.1 — Property of image sequences in NAIST campus package 02

Properties/ Seq #	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
Significant moving occluders										✓	✓	✓						
Fast camera movement													✓	✓				
Auto luminance control															✓	✓	✓	✓
Auto focus control																	✓	✓
Reference data	✓									✓			✓	✓				

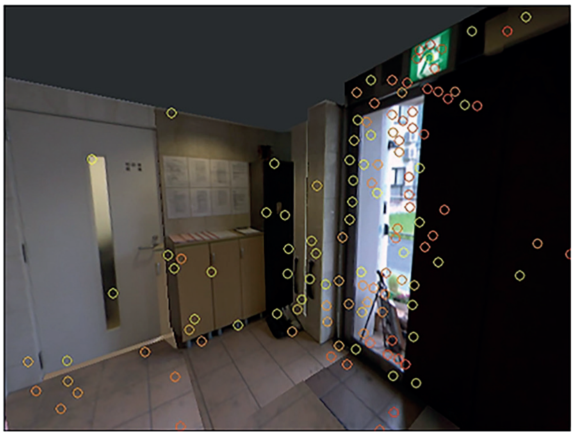
A tool has been developed to generate datasets for benchmarking using virtualized reality models. The aim of using virtualized reality models is to obtain ground truth data at low cost. In the case when virtualized reality models can be obtained, this tool can generate datasets that are composed of images generated from models and ground truth data. The ground truth data includes tracking data of interest points, which are 3D-2D correspondences of interest points, in addition to intrinsic and extrinsic camera parameters. As shown in Figure A.3, four open datasets generated from virtualized reality models have been provided. Moreover, 3D model data of the venue of ISMAR 2009, which is a material of the data shown in Figure A.3 a), has also been open to the public.



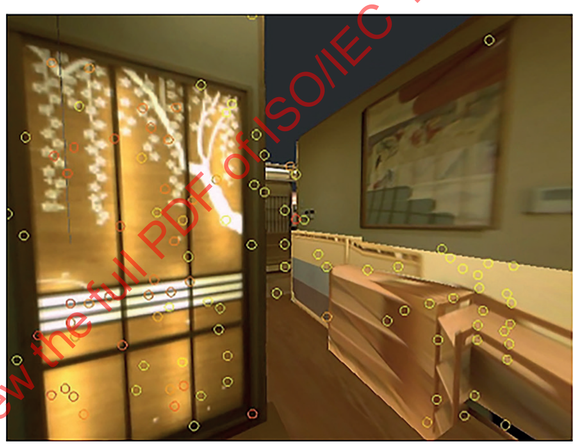
a) Venue of ISMAR2009



b) Tracking competition room for ISMAR2010



c) Nursing home



d) Japanese restaurant

Figure A.3 — Sample images of datasets generated from virtualized reality models

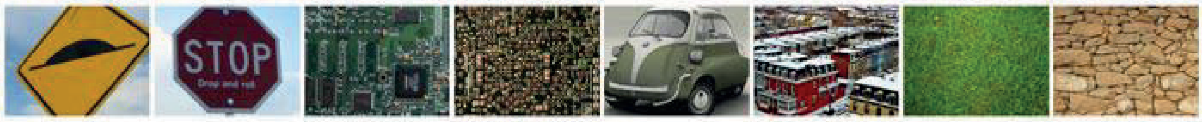
A.2 Dataset and evaluation for template-based tracking algorithms

In the case of a benchmarking activity for template-based tracking algorithms[7] conducted by a private sector, a benchmarking dataset including image and ground truth sequences was recorded by using a highly precise measurement arm mounted with an industrial camera (Figure A.4). Features of the dataset are realistic imaging conditions and very precise motions. When generating the dataset, the texture of the targets was carefully selected and the camera motions were as representative as possible.

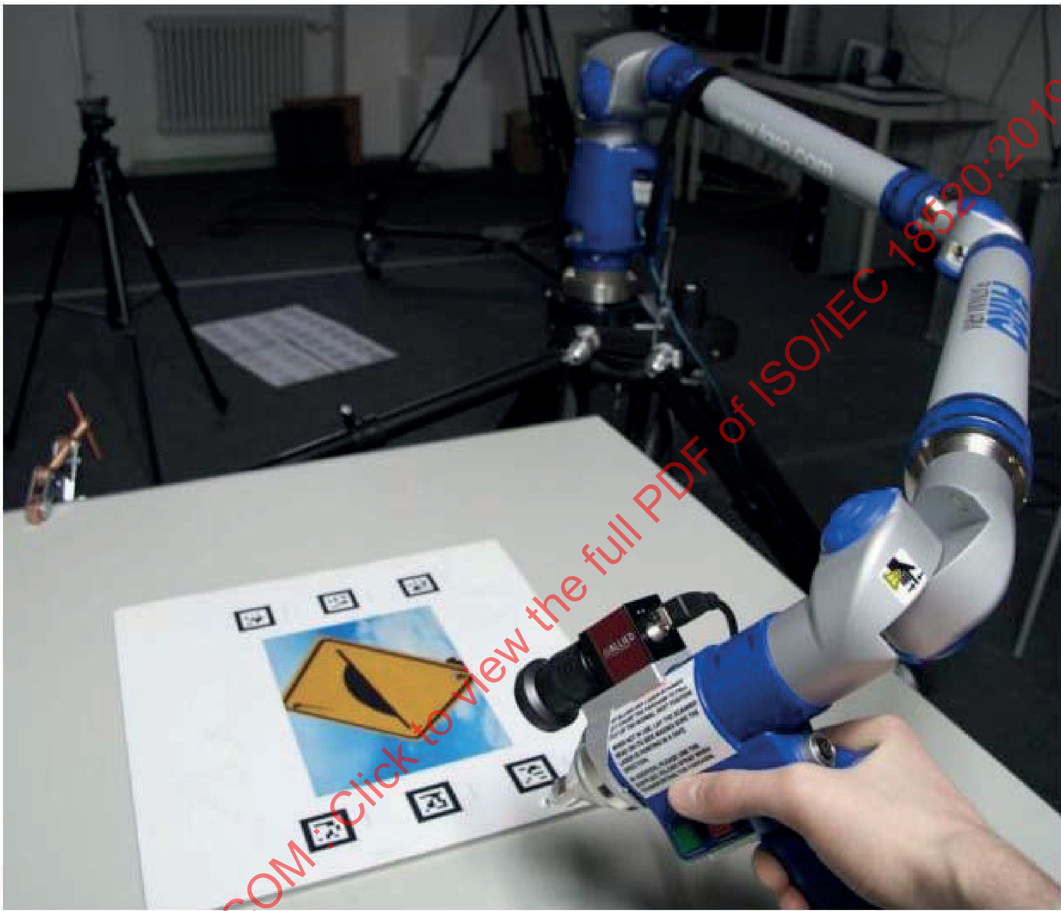
The evaluation is based on four reference points which are placed on the diagonal lines of the reference images (marked with blue crosses in Figure A.5), located at the XGA resolution boundaries, i.e. at (512;384). For every image $\mathbf{I}_i$ per sequence, the RMS distance err_i of each imaged reference point, $\boldsymbol{x}_j$, to the ground truth point, $\boldsymbol{x}^*_j$, is computed as

$$err_i = \sqrt{\frac{1}{4}\sum_{j=1}^{4}\left\| x_j - x_j^* \right\|^2}$$

After computing these errors for a sequence, all frames with an err_i of 10 px are removed as the cases with a higher RMS error as a sign that the tracking algorithm lost the target. Based on these filtered results, the ratio of tracked frames is computed and the distribution of the error is analyzed for the evaluation. Benchmarking service for template-based tracking algorithms had been provided by the private sector from 2009 to 2015, employing the datasets and the evaluation method mentioned above.



a) Reference targets used in the database
(from left: low, repetitive, normal, high texturedness)



b) The measurement arm, the camera and one of the targets

Figure A.4 — Datasets for template-based tracking algorithms

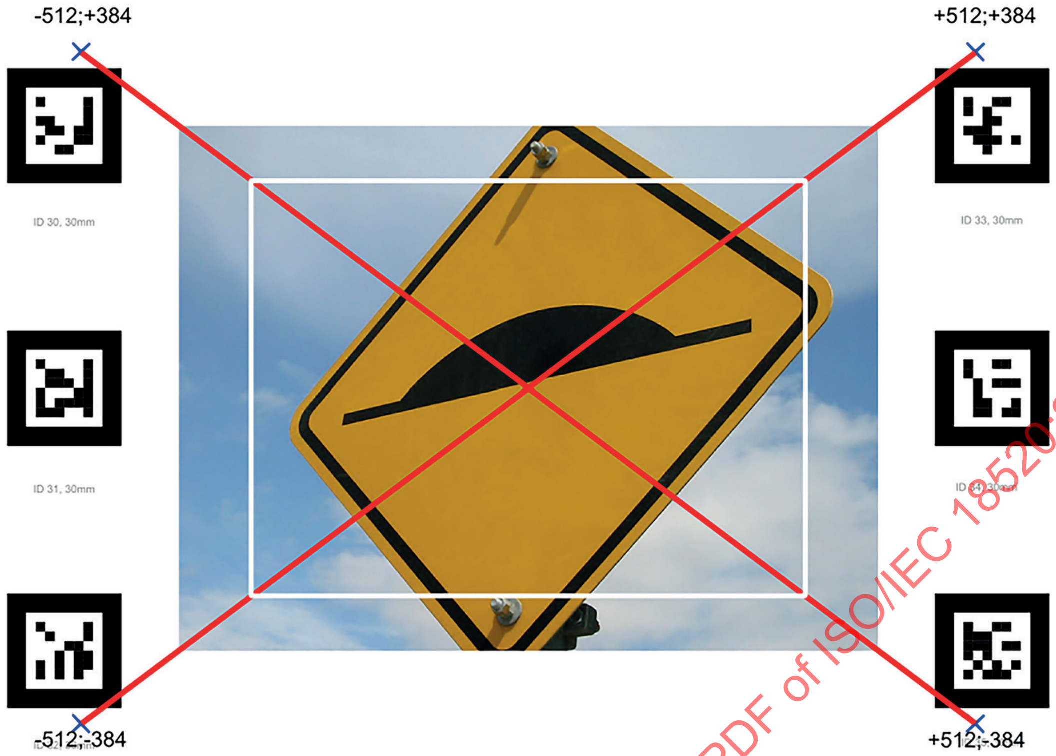


Figure A.5 — A reference target with six fiducials

A.3 Dataset for outdoor handheld camera localization

In the case of a benchmarking activity for outdoor handheld camera localization conducted by the same private sector as in A.2, a benchmarking dataset had been open to the public from 2013 to 2015. The dataset comprised a precise environment model and 125 image sequences of an urban environment captured under natural camera motions and different illumination settings (Figure A.6). For all these images, the dataset not only contained readings of the sensors attached to the camera such as GPS positions, compass headings and measurements of the gravity vector, but also ground truth information on the geometry and texture of the environment and the full 6DoF camera pose. The extensive procedure to create the outdoor 6DoF ground truth dataset is described in Reference [8].

Figure A.6 — – Dataset comprising a precise environment model and over 45 000 camera images with sensor values and 6DoF ground truth poses

A.4 Dataset and evaluation toolkit for mobile augmented reality

A.4.1 General

By HITLAB NZ, the source code of the mobile augmented reality visualization and evaluation toolkit and datasets to support reproducible research have been published (Figure A.7)[9].

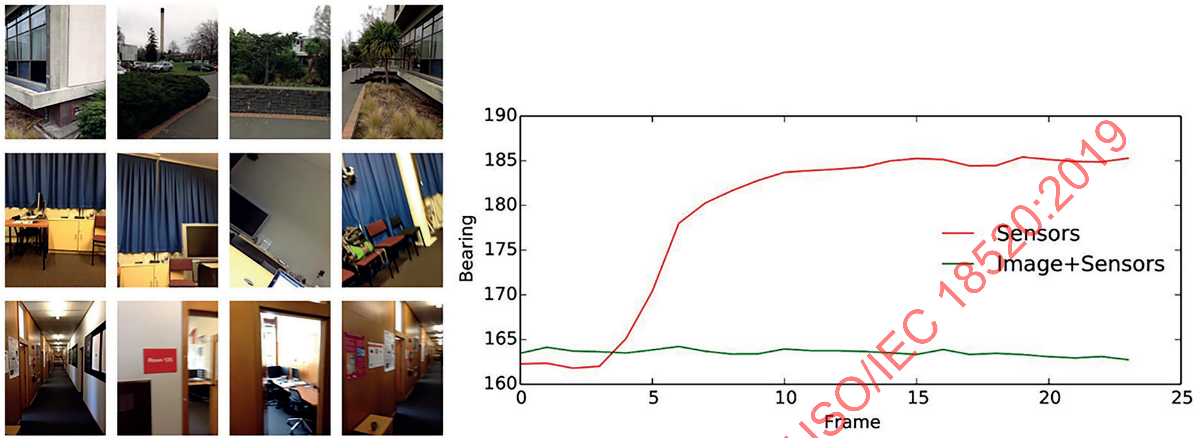


Figure A.7 — Examples of HITLAB NZ datasets (left) and example result on comparison between sensor-based tracking and sensor-fusion-based tracking (right)

Figure A.8 shows the overview of typical data flow in Transform Flow[9], which has been released under the MIT license on GitHub. It was originally developed as part of a single research project but is being released to the community to encourage collaborative research, development and education in the area of mobile AR.

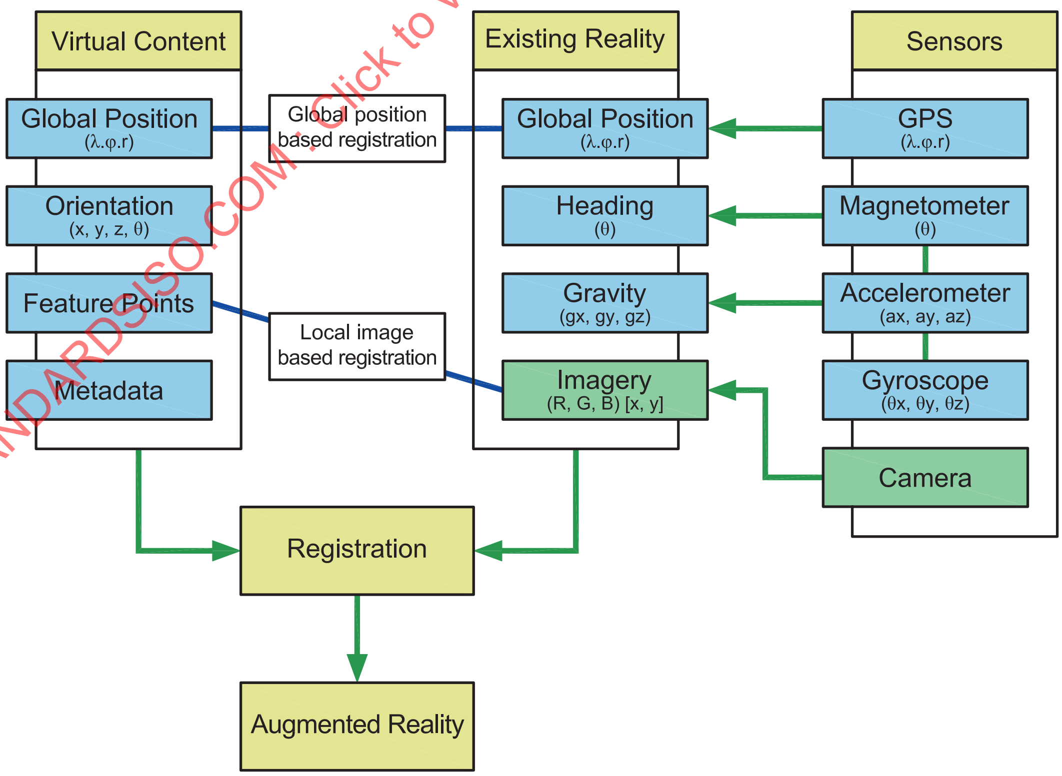


Figure A.8 — Overview of typical data flow in mobile outdoor AR

A.4.2 Data capture

A mobile data acquisition system (Figure A.9) was created to capture sensor data and video frames. This application was developed using Objective-C. The rate at which sensor data and video frames are captured is independent, and can be customized at compile time. All data is saved in a CSV (comma separated values) formatted log file, including video frames that reference external PNG (portable network graphics) image files in the same directory.

Sensor data is captured using frameworks in a smartphone. Core Location provides WGS84 latitude, longitude, altitude and bearing, while Core Motion provides gravity, linear acceleration rotation rate and magnetic flux. The position and bearing typically have high latency, while the motion data is usually captured at 30 Hz. Video frames are captured using AV Foundation, which provides access to the camera at a variety of resolutions. The captured image is recorded as interpolated RGB, which is universally supported across all devices and provides a convenient format for further processing. The system typically capture at 480 × 360 images at 10 Hz — higher frame rates generate very large amounts of data and can be more difficult to analyze.

Phone specific data including the device name and hardware identification are recorded as one of the first entries in the log file. This data can be used for hardware or device specific calibration files. In order to support global tracking algorithms, when recording is triggered, the current position and bearing are written to the log file if possible. This allows at least one global position and bearing entry to exist before frame data and motion data, which is typically required for initialization of the system.

The tool includes a real-time visualization of sensor measurements. This allows for a greater understanding of the effects that the physical device motion has on the sensor data. It can be instructive to check the gyroscope, accelerometer or other sensors while manipulating the device. Device-specific issues including calibration problems and drift can be assessed quickly and isolated.

Figure A.9 — The data capture application running on a smartphone

A.4.3 Visualization

A desktop application[9] was developed for viewing datasets captured using the mobile data acquisition tool (Figure A.10). This application is written in C++11 and uses OpenGL for rendering. Current implementation assumes that the user is interested in visualizing data within a global frame of reference, with position defined by (latitude, longitude, altitude) and rotation defined by (bearing, gravity). To systematically capture this data, an abstract motion model is defined to process incoming sensor data and image frames. The camera pose is computed in two steps, a quaternion rotation based on the bearing and gravity information, and a displacement based on the latitude and longitude. East, North, Up (ENU, maps to XYZ) Cartesian coordinates are used which are intuitive and practical for the small datasets that are usually being evaluated. The input dataset, combined with a motion model, produces an immutable per-frame globally registered camera pose. This sequence of frames can then be visualized and analyzed without further motion processing. In the case that a motion model requires several sensor updates before it can be initialized reliably, the visualization starts at the first frame after the localization becomes valid.

Displaying frames in a 3D environment is challenging because a camera frame is a planar projection of a typically non-planar environment. Without depth information, it is impossible to accurately reproduce the actual 3D structure. To work around this problem, the visualization tool uses planar projections based on the camera's field of view and an arbitrary scale factor. Frames can then be rendered in a 3D environment and explored using an arbitrary view position. Pure rotations are the easiest to interpret visualization as there is no change in planar alignment.

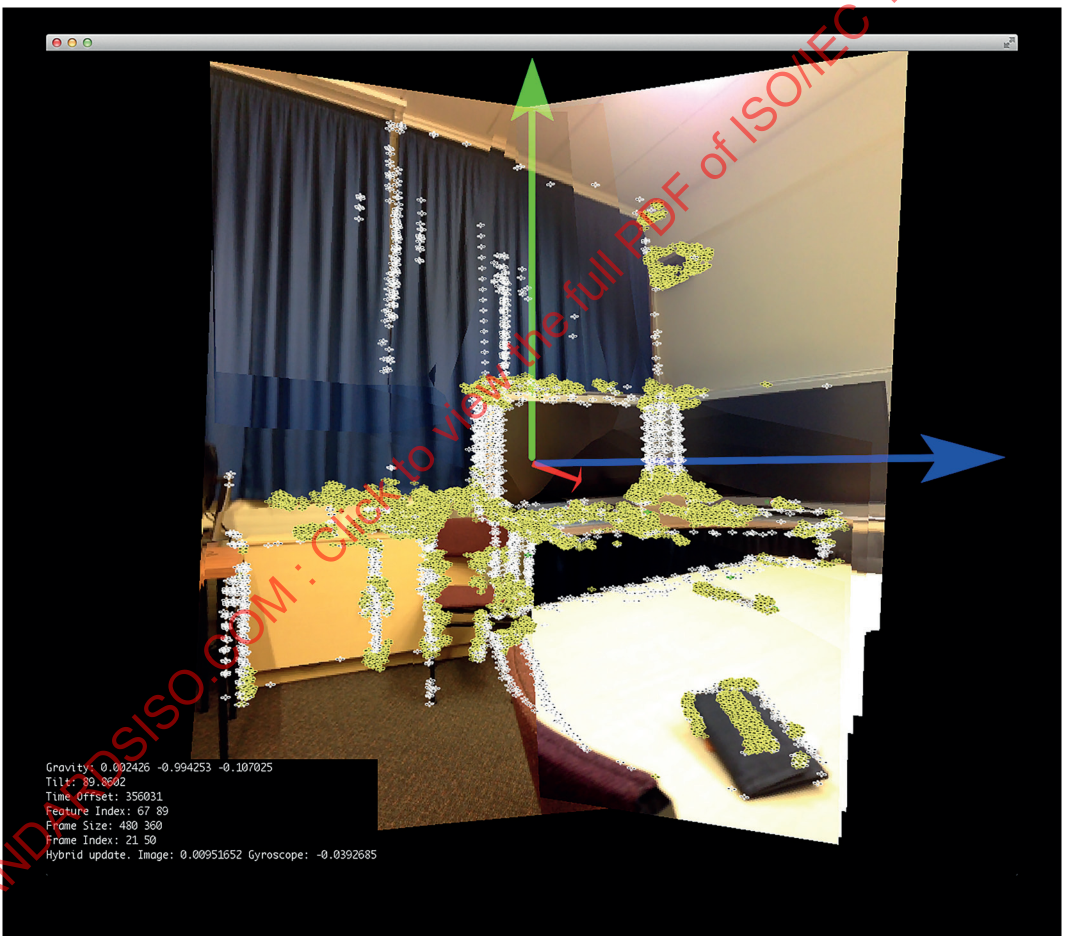


Figure A.10 — The Transform Flow Visualization application rendering 50 combined frames in a 3D environment

A.4.4 Deployment

An application has been developed, the Transform Flow Browser (Figure A.11), which can be used to run algorithms developed using the Transform Flow motion model abstraction. It uses a similar setup as the capture tool, but rather than logging the events, it applies them directly to a motion model. The AR visualization is then rendered using the calculated frame of reference.

The browser can display points of interest using 3D content (via Wavefront OBJ files) and 2D billboards (constructed directly from UIView instances), which are rendered on top of a real-time video stream using OpenGLES. The visualization includes a planar grid which is useful for understanding the practical tracking quality, e.g. jitter, rotations. The implementation itself is multi-threaded, and uses Grand Central Dispatch to offload the rendering, camera frame capture and motion model computations to separate CPU cores so that the main user interface remains responsive. Dealing with a wide variety of camera and screen configurations is not trivial. Specifically, different cameras have different intrinsic properties, the most important being the field of view.

Device specific calibrations can be useful in a research setting but would be cumbersome for end user applications. Ideally, a per-device database of camera intrinsic properties would allow for widespread deployment with acceptable accuracy, or for unknown devices, some type of online calibration procedure would suffice.

Figure A.11 — The Transform Flow Browser, showing a 3D model overlaying the HITLAB NZ image

A.5 Trial set — The City of Sights

A.5.1 General

The City of Sights[10] was designed and implemented as a physical and virtual model of an imaginary urban scene. It can serve as a backdrop or "stage" for a variety of AR research. It can be used for evaluation of such AR topics as tracking systems, modelling, spatial AR, rendering tests, collaborative AR and user interface design. By openly sharing the digital blueprints and assembly instructions for their models, their dataset was allowed to be physically replicable by anyone and be allow customization and experimental changes to the stage design which enable comprehensive exploration of algorithms and methods. Furthermore, an accompanying rich dataset consisting of video sequences under varying conditions with ground truth camera pose was provided. Three different ground truth acquisition methods were employed to support a broad range of use cases.

A.5.2 Model selection, design and construction

The “City of Sights” consists of models of the following buildings.

— The Pyramid of Cheops (also called the Pyramid of Khufu)

— The Berliner Dom (Berlin Cathedral)

— The Arc de Triomphe de l’Etoile in Paris

— The Musikverein in Vienna (Vienna concert hall)

— A medieval Irish Round Tower

— St. Mark’s Campanile in Venice

Objects of different geometric complexities, including different geometric primitives such as boxes, approximate cylinders (the Round Tower), domes (on the Berliner Dom), concave surfaces (most prominently in the Arc de Triomphe) and small details (the rims of St. Mark’s Campanile) as well as a variety of textures ranging from complex to repetitive to low textured are intentionally chosen.

The overall size and the number of objects are limited so that the model at its default scale fits on a ground plane of 800 × 550 mm (slightly smaller than A1), and that the total time needed for construction remains manageable. The models are arranged so that measurements of unobstructed views of all buildings and views with occlusions are possible with the limited reach of a robot arm. The object sizes were chosen such that all paper folding plans fit on A3 or 11 × 17" sheets. The absolute scales of the models are not matched; the ease of printing and assembly was considered to be more important.

A.5.3 Acquisition of video and ground truth data

Three different methods were used for camera control and ground truth acquisition: a robot arm, a manually guided and mechanically tracked coordinate measuring machine (CMM) arm and an optical tracking system. All measurements used in the various calibration steps were recorded and were available together with the video and ground-truth data (Figures A.12, A.13 and A.14). This enabled every researcher to both assess the quality of the sensor data and employ different algorithms if desired.

— **Robot arm sequences:**

The computer-controlled robot arm was used to make “perfect” image sequences without the problems of handheld camera movements. It has the advantage that the exact same camera path can be repeated while changing properties of the environment such as models, lighting and camera parameters. The robot arm’s position was moved in 0,1 mm increments between frames, and still images of 1 600 × 1 200 pixel resolution were captured free of motion blur, frame drops or jitter. The resulting sequences can then be combined into a video sequence or treated as separate high-quality images.

— **Mechanically and optically tracked sequences:**

Mechanical and an optical tracking setup were used to provide additional image data with ground truth. The ground truth is less precise than for the robot arm, but the setup allows for direct user-controlled camera movements including real-world camera issues such as motion blur or jitter.

Dataset	Video	Individual frames	Reference data	Description
Light condition 0				
CS_RA_L0_BirdsView:	.avi (12 MB)	.zip (359 MB)	.csv	bird's view, 2088 frames
CS_RA_L0_TopView:	.avi (10 MB)	.zip (260 MB)	.csv	top view, 1809 frames
CS_RA_L0_StreetView:	.avi (14 MB)	.zip (367 MB)	.csv	street level view, 2269 frames
Light condition 1				
CS_RA_L1_Groundplane:	.zip (?? MB)	.zip (344 MB)	.csv	groundplane only, 1808 frames
CS_RA_L1_BirdsView:	.zip (?? MB)	.zip (319 MB)	.csv	bird's view, 2088 frames
CS_RA_L1_TopView:	.zip (?? MB)	.zip (310 MB)	.csv	top view, 1809 frames
CS_RA_L1_StreetView:	.zip (?? MB)	.zip (367 MB)	.csv	street level view, 2270 frames
Light condition 2				
CS_RA_L2_BirdsView:	.zip (?? MB)	.zip (363 MB)	.csv	bird's view, 2090 frames
CS_RA_L2_TopView:	.zip (?? MB)	.zip (219 MB)	.csv	top view, 1000 frames
CS_RA_L2_StreetView:	.zip (?? MB)	.zip (363 MB)	.csv	street level view, 2269 frames

Figure A.12 — Robotic arm sequence dataset

Dataset	Video	Individual frames	Reference data	Description
CS_ART_02:	(.zip, XX MB)	.zip (72 MB)	.csv	groundplane only
CS_ART_02_augm:	(.zip, XX MB)	.zip (95 MB)	.csv	as above, with augmentation
CS_ART_03:	.avi (8 MB)	.zip (85 MB)	.csv	entire scene
CS_ART_04:	.avi (11 MB)	.zip (119 MB)	.csv	entire scene
CS_ART_05:	.avi (11 MB)	.zip (119 MB)	.csv	entire scene
CS_ART_06:	.avi (12 MB)	.zip (119 MB)	.csv	entire scene
CS_ART_07:	.avi (11 MB)	.zip (119 MB)	.csv	entire scene
CS_ART_08:	.avi (14 MB)	.zip (119 MB)	.csv	entire scene
CS_ART_09:	.avi (11 MB)	.zip (119 MB)	.csv	entire scene
CS_ART_10:	.avi (12 MB)	.zip (119 MB)	.csv	entire scene
CS_ART_11:	.avi (12 MB)	.zip (119 MB)	.csv	entire scene
CS_ART_12:	.avi (12 MB)	.zip (119 MB)	.csv	entire scene

Figure A.13 — Optical tracking sequence dataset

Dataset	Video	Individual frames	Reference data	Description
CS_FARO_02:	.avi (9 MB)	.zip (32 MB)	.csv	ground plane only, 1101 frames
CS_FARO_02_augm:	.zip (?? MB)	.zip (145 MB)		as above, with augmentation
CS_FARO_03:	.avi (XX MB)	.zip (78 MB)	.csv	entire scene
CS_FARO_04:	.avi (9 MB)	.zip (119 MB)	.csv	entire scene
CS_FARO_05:	.avi (9 MB)	.zip (119 MB)	.csv	entire scene
CS_FARO_06:	.avi (11 MB)	.zip (119 MB)	.csv	entire scene
CS_FARO_07:	.avi (13 MB)	.zip (119 MB)	.csv	entire scene
CS_FARO_08:	.avi (14 MB)	.zip (119 MB)	.csv	entire scene
CS_FARO_09:	.avi (14 MB)	.zip (119 MB)	.csv	entire scene
CS_FARO_10:	.avi (13 MB)	.zip (119 MB)	.csv	entire scene
CS_FARO_11:	.avi (12 MB)	.zip (119 MB)	.csv	entire scene
CS_FARO_12:	.avi (11 MB)	.zip (119 MB)	.csv	entire scene

Figure A.14 — CMM arm sequence dataset

A.6 Tracking Competition in ISMAR 2014

A.6.1 General

Tracking competitions have been organized in ISMAR conferences almost every year since 2008[11]. In this subclause, a tracking competition in ISMAR 2014[12] is briefly introduced. The aim of the competition was to evaluate state-of-the-art tracking approaches in industrial settings. The competition consisted of four scenarios: marker-less tracking of a rotating miniature model, tracking and learning on different vehicles, tracking with high accuracy and tracking inside an unknown area.

A.6.2 Scenario 1 — Tracking of a rotating vehicle

The task for Scenario 1 was to localize 3D coordinates on a rotating vehicle model. In other words, the tracking algorithm provided correct superimposition of virtual data on the real vehicle (see Figure A.15).

In the preparation phase, participants had to register/adapt to the local vehicle coordinate system using uncoded circular markers whose exact coordinates were provided or were known 3D data. In the competition phase, points on the base underneath the vehicle had to be identified using predefined 3D coordinates. Then 3D data had to be overlaid on top of the rotating vehicle model as accurately as possible. The reference points previously used for calibrating were no longer available on the rotating vehicle base (oval shaped disc). Moreover, tracking was permitted only in a limited area. Therefore, the difficult challenge of this scenario was to track a moving object, the vehicle, with variable speed from a fixed position.

In the first three sub-tasks, a point was awarded by the jury when an element defined by 3D coordinates was correctly identified. This point was multiplied with the previously established level of difficulty, which was determined by the speed and variance of the rotation. In addition, the jury evaluated the overlay quality of each approach. Decisive criteria were robustness, accuracy, the initializing procedure and duration of the overlay.

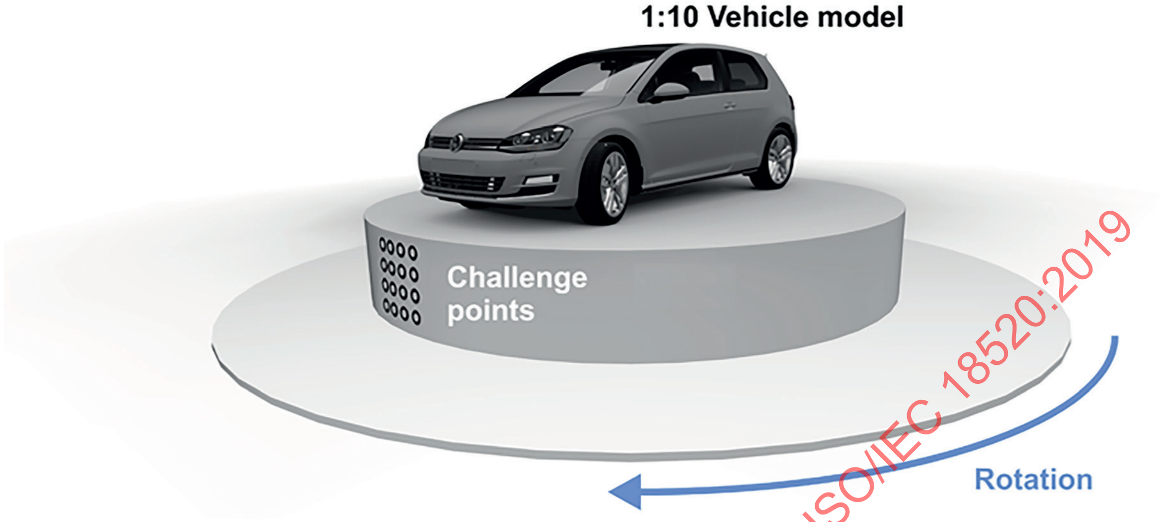


Figure A.15 — Tracking target of Scenario 1

A.6.3 Scenario 2 — Tracking and learning on different vehicles

The task for Scenario 2 was to identify exterior and interior parts of a vehicle defined by 3D coordinates.

In the registration phase, participants captured the external appearance of the complete vehicle. Then, they registered it with the help of uncoded circular markers in the engine compartment or 3D data of parts of the exterior and interior in the local coordinate system of the vehicle. The exact coordinates of the markers were provided beforehand. During the competition, the task was to identify specific items based on given 3D coordinates in four different areas of the vehicle (interior and exterior) in a fixed order (see Figure A.16). Before starting the competition phase, the reference points previously used for registration to the local coordinate system were removed. In addition, the appearance of the vehicle changed: lighting was varied and vehicle elements were removed or added. Moreover, in the interior as well as in some areas outside the engine compartment, there was no information about the local coordinate system.

In each case, a point was awarded for the correct assignment of an element to a given 3D coordinate. This point was multiplied with the predetermined level of difficulty of each subtask. Furthermore, the time needed to perform the competition task was measured, and the time affected the evaluation when multiple teams scored the same number of points.

Figure A.16 — Target areas of Scenario 2

A.6.4 Scenario 3 — Tracking with high accuracy

The task for Scenario 3 was to place adapters accurately at positions defined by 3D coordinates.

Unlike other scenarios, in this scenario the preparation and competition phase directly flowed from one to the other. The participants were allowed to place their own markers and features into a specified area in the preparation phase (see Figure A.17). After this, these markers and features had to be registered to the local coordinate system which was defined by circular reference points. The exact 3D coordinates of the reference points were provided. Before the start of the competition phase, the circular reference points were removed. In the competition phase, tracking was performed exclusively using the self-created markers and features.

Points were awarded according to the accuracy and speed of the tracking approaches. The accuracy was measured by the 3D error between the estimated position and the ground truth.

A.6.5 Scenario 4 — Tracking inside an unknown area

The task for Scenario 4 was to track in an unknown corridor and identify certain elements based on given 3D coordinates (see Figure A.18).

The participants were provided with a local coordinate system at the starting point which was defined by circular reference points. The starting point was in front of an unknown and winding corridor. The exact 3D coordinates of the reference points were known. During the competition, the task was to move through the unknown corridor and identify certain elements. The participants were not able to see the corridor in the preparation phase. The corridor was visible only during the competition phase. It was only known in advance that the corridor would be about 20 m long.

For each of the challenge points, rating points were given as follows: a rating point for correctly labelled challenge points at the start, two rating points for correctly labelled challenge points in the middle of the corridor and four rating points for correctly labelled challenge points at the end.

A.6.6 Evaluation method

Figure A.19 and Table A.2 show how to evaluate the result of each scenario. In Scenarios 1, 2 and 4, the jury judged whether challenge points were correct or not. On the other hand, the 3D error between the estimated point and the ground truth was used as the evaluation criterion for Scenario 3.

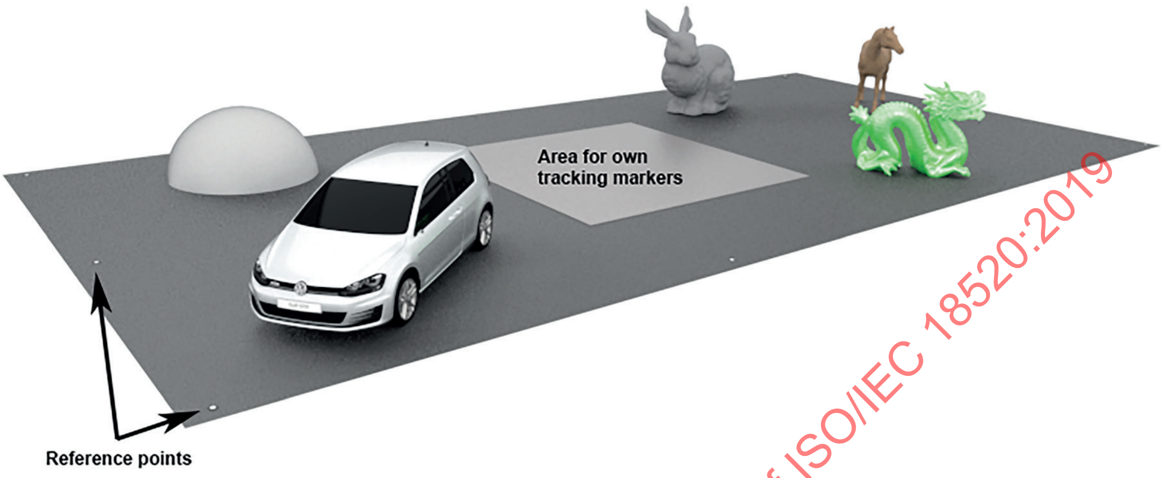


Figure A.17 — Tracking area of Scenario 3

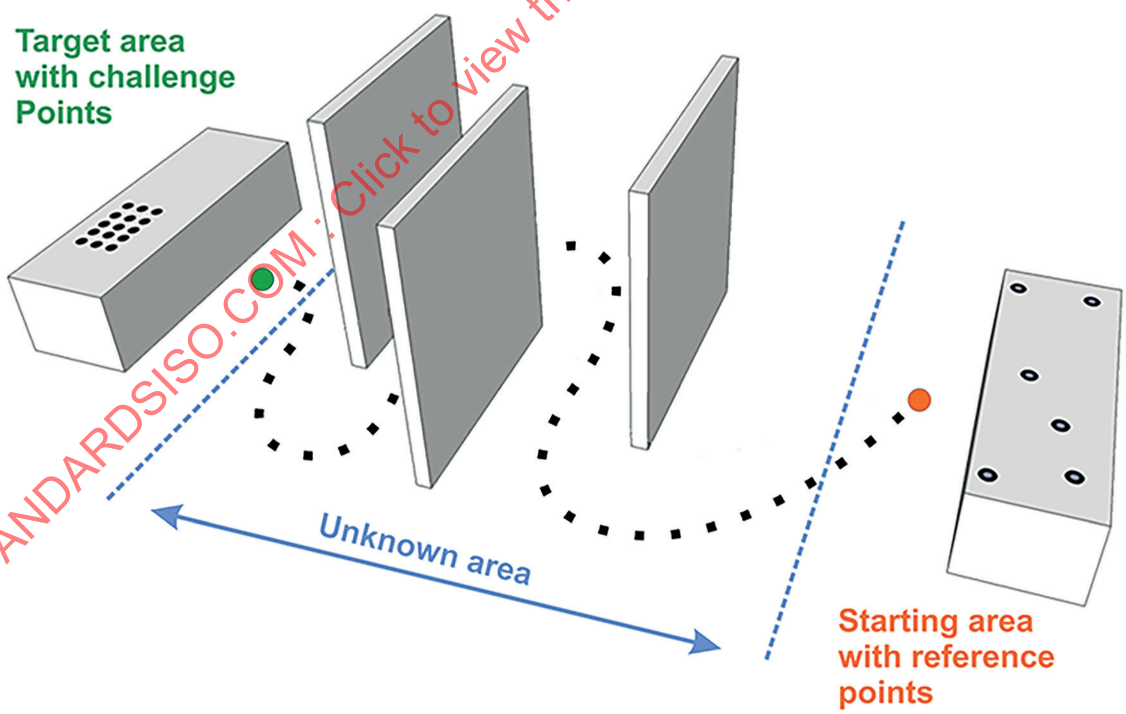


Figure A.18 — Tracking environment of Scenario 4

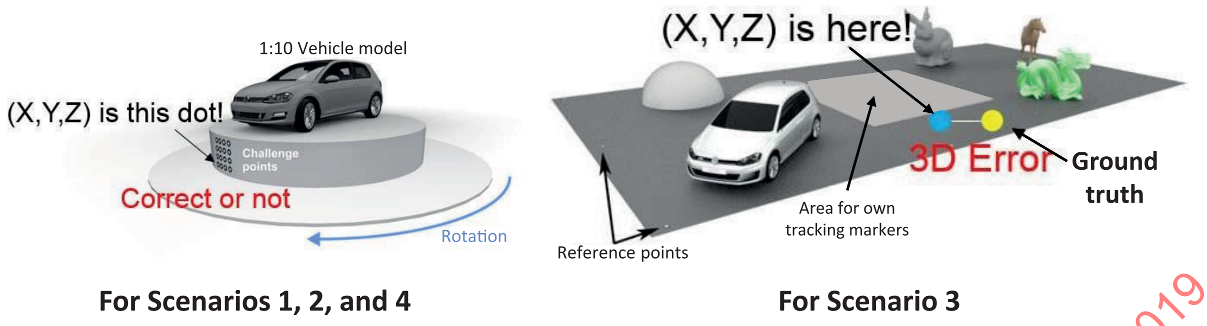


Figure A.19 — Evaluation methods for each scenario

Table A.2 — Objective and subjective indices

Scenario	Objective	Subjective
Scenario 1	Whether challenge points are correct or not	Visual quality of rendered virtual objects
Scenario 2	Whether challenge points are correct or not	N/A
Scenario 3	3D error	N/A
Scenario 4	Whether challenge points are correct or not	N/A

A.7 Tracking Competition in ISMAR 2015

A.7.1 General

This subclause presents the tracking competition in ISMAR 2015[13].

IEICE PRMU (Special Interest Group on Pattern Recognition and Media Understanding in the Institute of Electronics, Information and Communication Engineers) algorithm contest committee and the ISMAR 2015 Tracking Competition committee including TrakMark WG collaborated in organizing the tracking competition in ISMAR 2015. The ISMAR 2015 tracking competition was composed of an off-site tracking competition and on-site tracking competition. Especially, the off-site competition was the first attempted by ISMAR.

A.7.2 Off-site tracking competition

The off-site tracking competition was designed for students and non-experts to encourage them to participate in tracking competitions. The process of camera pose tracking was divided into three steps: point matching, point tracking and mapping (see Figure A.20). The levels of the off-site tracking competition corresponded to each step. In the competition, participants downloaded a dataset and submitted the tracking result for each level. Through the competition, participants learned one of the techniques for tracking camera poses in augmented reality. The participants acquired enough knowledge to attend the on-site tracking competition by becoming knowledgeable in all levels of the off-site tracking competition. The tasks of the off-site competition were as follows.

— Level 1: Finding the locations of known feature points in an image.

— Level 2: Tracking known feature points in successive frames.

— Level 3: Tracking known feature points and mapping new feature points in successive frames.

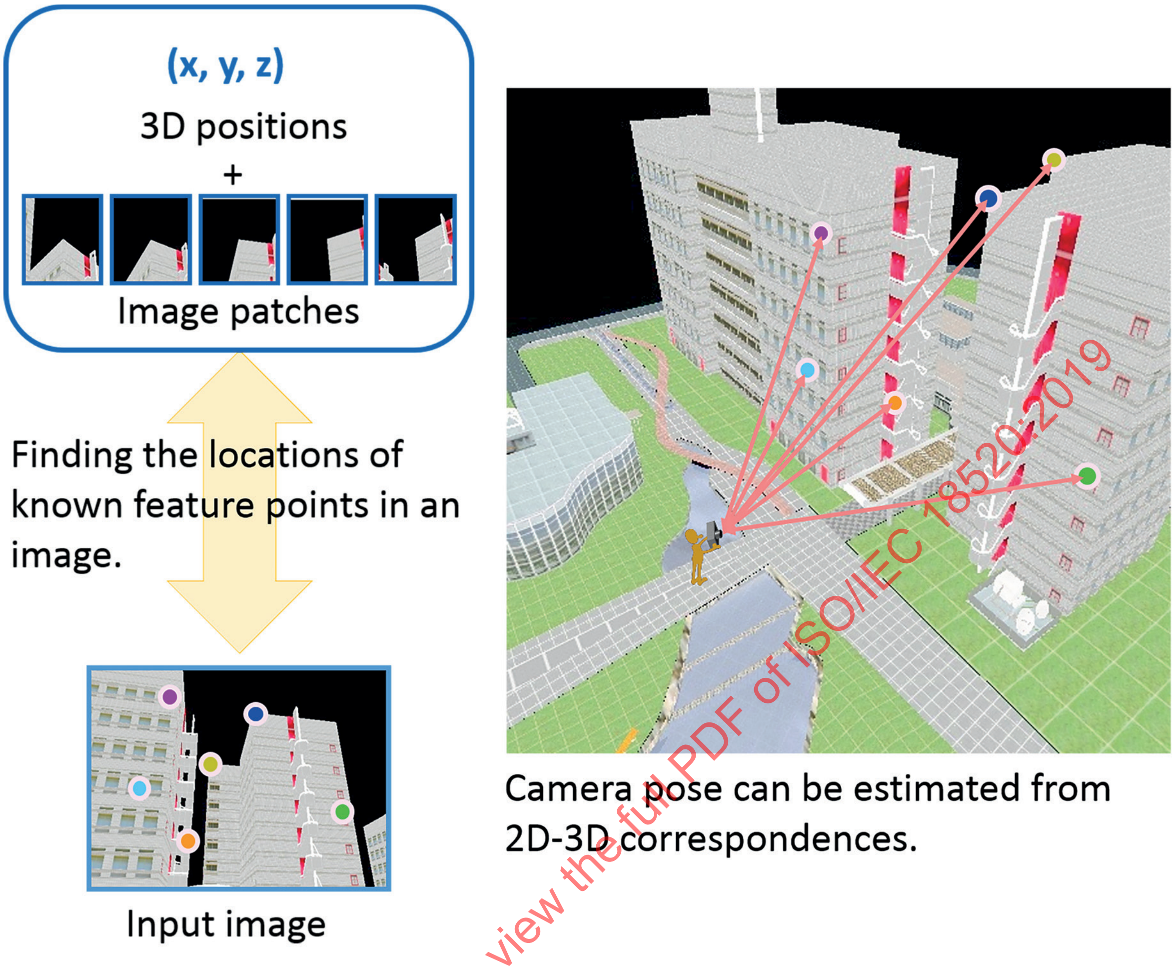


Figure A.20 — Overview of the off-site competition

Template source codes for Level 1 and Level 2, which included a simple template matching method with the C++ standard template library and OpenCV, was provided. Participants of Level 1 and Level 2 had to use the template codes and could modify ONLY `user_function()` in user_function.cpp to implement their original ideas. On the other hand, no sample codes for Level 3 were provided. Participants of Level 3 could not only revise all the template codes for Level 1 and Level 2, but also implement codes from scratch.

Synthetic and real scenes from Level 1 to Level 3 were used for all scenarios. As described later, the tracking scenes were constructed using plastic building blocks. Input images were prepared by generating synthesized images using POV-Ray, which is a tool for rendering high-quality computer graphics for synthetic scenes and capturing still images using a camera mounted on a robot arm for real scenes. The resolution of the images was 640 × 480 pixels. The target scene was composed of rigid objects and illuminated by multiple static or dynamic light sources.

The scenes were constructed using plastic building blocks from the following viewpoint.

— Feasibility:Plastic building blocks were sold worldwide and are easy to obtain and construct.

— Reliability: Accuracy was guaranteed regarding assembly.

— Relevance: Image features were obtained, and moderate specular reflection was observed.

For Scenarios 1 and 3, a 3DCG model was created using a 3D modeling tool which enables a user to build models using virtual blocks and converted to the POV-Ray format using LDraw. Camera movement was designed and exported to the POV-Ray format using a Java program. For Scenario 1 and 2, a purchasable model was used.

Scenarios

Scenario 1

For each level, zipped PNG images and zipped patch data were provided. For Level 3, intrinsic camera parameters were provided for mapping.

Scenario 2

The scene of this scenario was the same block house, but a real one. All images were taken by a real camera. The 3D world coordinate system and its origin were also the same as Scenario 1. That is, hand-eye calibration was performed, and camera poses obtained from the robot arm were used as ground truth.

Scenario 3

The target scene of this scenario was a block world placed on a flat surface. This world was an extension of the block house used in Scenario 1. All images were generated by POV-Ray. The origin of the 3D world coordinate system was the same in Scenario 1.

A.7.3 Off-site competition procedure

Level 1

The detailed instructions of the Level 1 off-site tracking competition were as follows:

1. Determine the 2D positions of the reference points given by 2D image patches extracted from the image I_0.
2. Calculate the projective transformation matrix referring to the correspondences between the 2D positions of the reference points in the image I_0 and their 3D coordinates.

In general, a camera pose is determined from more than three correspondences between the 2D positions of scene points and their 3D coordinates by solving the perspective-n-point (PnP) problem. The Level 1 algorithm has to return the projective transformation matrix of image I_0.

Level 2

The detailed instructions of the Level 2 off-site tracking competition were as follows:

1. Determine the 2D positions of the reference points given by 2D image patches in the initial image I_0 of an image sequence.
2. Track them through the images I_0, ..., I_{n-1} in the sequence.
3. Calculate the projective transformation matrix referring to the correspondences between the 2D positions of the reference points in image I_{n-1} and their 3D coordinates.

All the given images of an image sequence contained the same scene part that could be found in a section of the initial image I_0. The Level 2 algorithm could use inter-frame restrictions of the given video sequence. In contrast to the Level 1 algorithm, the Level 2 algorithm has to return the projective transformation matrix of the image I_{n-1}. The participants could assume that the scene was rigid, the camera poses were consecutive, and the intrinsic parameters were not changed. However, the lighting environment could gradually change during camera movement.

Level 3

The detailed instruction for the Level 3 off-site tracking competition was as follows:

Calculate the projective transformation matrix referring to the correspondences between the 2D positions in the last image I_{n-1} of an image sequence and their 3D coordinates.

This task was similar to Level 2. However, any scene parts (patches) found in the initial image I_0 did not appear any more in the ending sequence of the images, including I_{n-1}. For Level 3, tracking only the given points might not be enough to achieve accurate estimation of the camera pose.

A.7.4 On-site tracking competition

The on-site tracking competition was held during the main conference in ISMAR 2015 (from Sep. 30 to Oct. 2). In the competition, tracking accuracy of visual simultaneous localization and mapping (vSLAM) techniques was evaluated, which was similar to Scenario 4 of the ISMAR 2014 Tracking Competition.

The competition environment was composed of a starting area and multiple challenge areas. The chessboard marker included in OpenCV 3.0.0 was placed at the starting area. The chessboard marker was printed on A0 paper and the size of a square was 100 mm. The world coordinate system was calibrated with a Total Station. Units of the world coordinate system were in mm.

A.7.5 On-site competition procedure

The detailed instructions of the on-site tracking competition were as follows (see Figure A.21):

1. Receive the 3D coordinates of challenge points, which are provided by USB memory, from the jury, and copy them to the vSLAM system.
2. Execute the vSLAM system on a mobile device and register the initial position using the world coordinate system specified by the chessboard marker.
3. Move the mobile device to find the challenge points in the specified order and mark them using a marker pen.

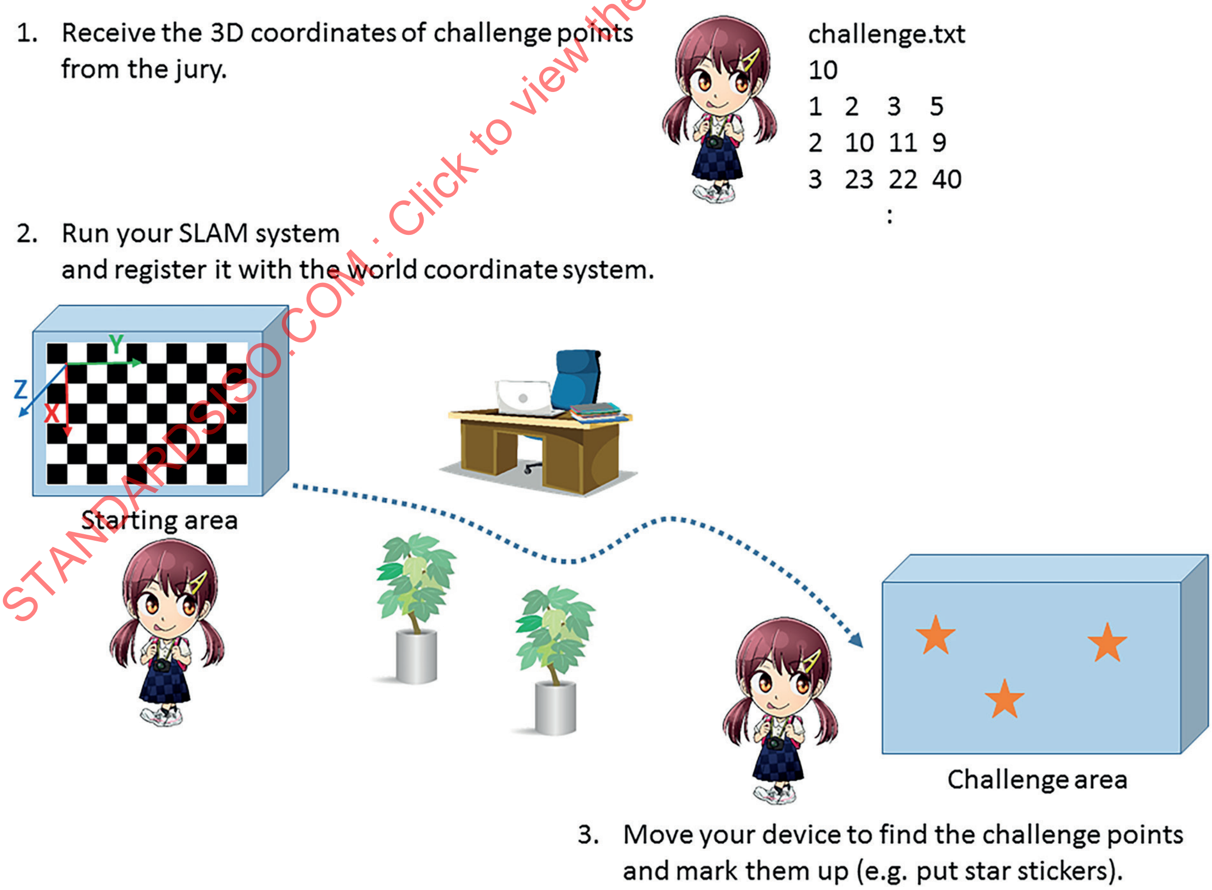


Figure A.21 — On-site competition

A.7.6 Device requirements and rules for the on-site competition

To carry out the competition fairly, the following device requirements were stipulated.

1) The device was equipped with one or more displays to visualize the annotations of challenge points. The device had to send a live video feed with annotations through an HDMI or mini-HDMI port to display the competitor's performance to the jury and audience.

2) The device was mobile and stand-alone. During the competition, one of the participants had to keep wearing, lifting or shouldering the system without placing it on the floor.

3) Any cables connecting to the environment, even for power supply, were not available.

4) Any infrastructure such as additional markers and Wi-Fi access points could not be installed in the environment. On the other hand, any number of cameras, inertial sensors and depth sensors was allowed to be used.

The participants had to obey the following rules.

1) The annotations of challenge points had to be visualized on the display in the specified order.

2) Only the current challenge point had to be visualized at any one time.

3) The annotations had to be simple 2D or 3D objects.

4) When the current challenge point was outside the user's field of vision, a type of navigation instruction could be visualized.

5) Challenge points had to be marked only where their annotations were visualized on the display, not by guesswork.

6) Challenge points could be skipped. The skipped challenge points could not be identified afterwards.

7) If the tracking process was aborted, a re-initialization was allowed, although the time count would continue to run.

8) The location of the chessboard marker was visible only at the starting area.

The order of participants was determined by a lottery called "amidakuji." Participants were allowed to attend other participants' competitions. It could not be guaranteed that lighting and background conditions were the same for all the participants.

A.7.7 On-site and off-site competition results

The results of the competition are available in Reference [14].

A.8 Findings from tracking competitions

A.8.1 General

This subclause presents a summary of findings[15] from benchmarking activities, especially from ISMAR 2015 tracking competition described in A.7.

A.8.2 Findings from the off-site competition

A.8.2.1 Benchmark indicator

PEVO is one of the most direct and intuitive indicators for vision-based geometric registration and tracking methods for MAR. PEVO works well, as it is sensitive to the error of the position and orientation of the camera. Actually, there were some cases in the competition in which PEVO was large even if re-projection error of image features was small.

The organizers had asked each contestant to submit the projective camera matrices for data with lens distortion. However, some contestants submitted normalized camera matrices which consist of position and orientation and do not contain the intrinsic parameters of the camera, and other contestants the projective camera matrices. This made it difficult for the organizers to fairly and equally evaluate each contestant.

A.8.2.2 Scalability

It is becoming more difficult to deal with various submissions owing to the increasing number of contestants. It is also difficult to make each contestant thoroughly understand the rules and regulations. It seems more appropriate that the organizers provide software for evaluation and the datasets and that each contestant submits only specific components.

A.8.2.3 Cheating and fine tuning

It is better for judges to evaluate each method themselves; for instance, by making each contestant submit the binary (executable) program. Otherwise, it is difficult to identify cheating such as by using future information for performing global optimization based on the entire data. In order to address fine tuning issues, it is effective to provide various types of datasets to contestants.

A.8.3 Findings from the on-site competition

A.8.3.1 Effects of simplification

Simplification of the benchmarking process is often necessary for practical operation of a competition, but the pros and cons have to be considered as follows.

[Equality and simplification]

In the competition, the conditions for each contestant were not strictly equal. Each contestant was supposed to mark a textured paper attached to a partition wall to indicate the specified challenge point, which is the simplest virtual object for accuracy evaluation estimated by an MAR system (see Figure A.22). For time-saving, the organizers did not change the paper for following contestants after a contestant marked the textured paper. Thus, the paper with the mark became a part of the environment for the following contestants. This might not be fair because the appearance of the paper changed due to the mark, even if the size of the mark was small.

Figure A.22 — Textured paper attached on a partition wall

[Measurement and simplification]

Ideally, the measurement for accuracy evaluation has to be based on the distance between the 3D coordinates of the challenge point given by the organizers and the 3D coordinates of the challenge point estimated by the contestant's MAR system in 3D space. However, evaluation was based on the 2D coordinates on the textured paper. This is because it is quite difficult to measure the distance of arbitrary points in the air in 3D space. More than three challenge points on a plane are necessary if the organizers want to strictly evaluate the 3D position and orientation of the camera with such 2D coordinates.

A.8.3.2 Benchmark indicators

In the on-site competition, the organizers adopted the following strategy for evaluation:

1) Compare the number of challenge points each contestant had.

If (1) is the same for several contestants,

2) Compare the mean distance, which corresponds to 3D errors of virtual objects, or 3DEVO for short.

If (2) is also the same for several contestants,

3) Compare times for trial completion.

In fact, only 1) was relevant because only one contestant completed the trial. In terms of comprehensive evaluation, it would be valuable to utilize other benchmark indicators such as temporality indicators including frame rate and latency.

A.8.3.3 Measurement process

From the organizer's point of view, on-site competitions are supposed to be held at various locations, so it would be better to be able to use the same or standardized tools in the preparation phase for measuring the environments and ground truth, and for determining the correspondence between the real world coordinate system and the local one. In the trial phase of this competition, the 3DEVO was measured manually with a ruler. It was easily realized because a challenge point is supposed to be located on a plane. However, if it is not on a plane, the right way of measuring the 3DEVO has to be considered.

From the contestant's point of view, there were two choices on acquiring/constructing environmental 3D models in each MAR system by each contestant: beforehand or during the trial. In this competition, online acquisition, which is visual SLAM in short, was required. Considering various scenarios for MAR system usage, it would be better to also encompass modelling before trials in a competition.

A.8.3.4 Human factor

Because the user of the MAR system was supposed to mark the paper attached to the partition wall while carrying the system, the camera approaching the paper often made image registration and tracking unstable (see Figure A.23). As expected, one of the key techniques for obtaining high scores in the competition was to devise and master the camera movement.

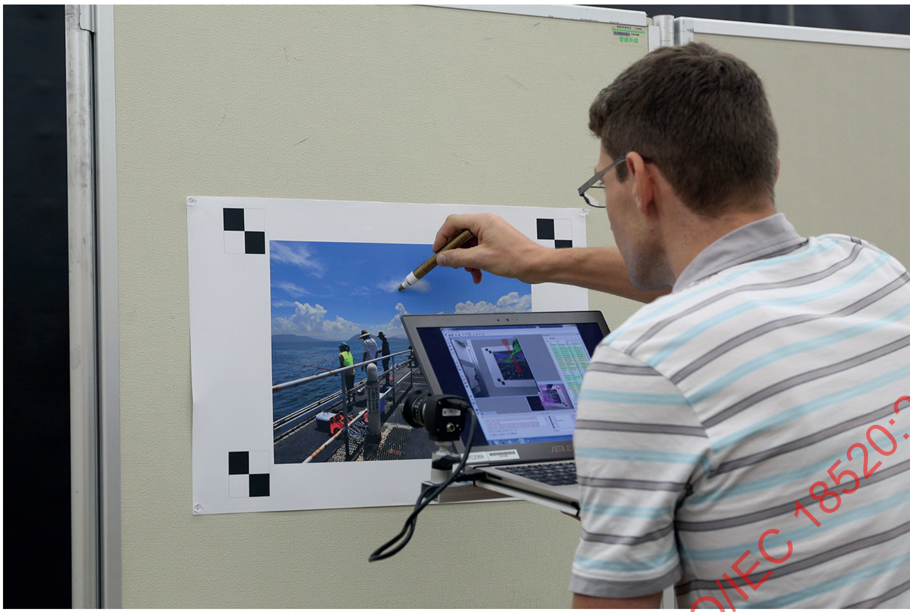


Figure A.23 — One of the contestants marking the textured paper

Even though the details of rules and regulations were documented, they were not well understood by the contestants because of the very busy schedule and it became problematic during the tracking competition. Visual aids such as presentation slides and handouts would help reconfirm the rules and regulations. Step-by-step confirmation during each test would be also helpful.

A.8.3.5 Difficulty level design

The difficulty of each trial of online registration and tracking strongly depends on the combination of objects aligned in the competition environment and their positions, so it is difficult to adjust the difficulty. However, plural challenge points alleviated the problem to some extent. In this competition, millimetric accuracy had some meaning for the first challenge point. As mentioned above, only one contestant completed the trial. For the following challenge points that involved 10 to 20 m movement, millimetric evaluation did not matter, and the goal was to keep tracking in stable fashion.

A.8.3.6 Dissemination

By showing screen-shot video of each MAR system during each trial on a large screen, the audience enjoyed the competition (see Figure A.24). Such efforts by the organizers were also effective in making the competition open and public.

Figure A.24 — The on-site competition venue

Annex B
(informative)

Usage examples of conformance checklists

This annex shows examples of how to use the conformance checklist shown in Table 3. Case examples of on- and off-site grass-roots benchmarking activities including competitions are introduced in A.1 to A.8. The following usage examples of the conformance checklist from Tables B.1 to B.12 correspond to the case examples.

Table B.1 — Conformance checklist for benchmarking activities in A.1

		Check	Item	Remarks
Process flow	Process	[✓]	Develop vSRT methods and/or MAR systems:	TrakMark committee members
		[✓]	Gather vSRT methods and/or MAR systems:	TrakMark committee members
		[✓]	Prepare and conduct benchmarking:	TrakMark committee members
		[✓]	Provide and maintain benchmarking instruments:	TrakMark committee members
		[✓]	Provide and maintain benchmarking repositories:	TrakMark committee members
		[✓]	Share benchmarking results:	TrakMark committee members
		Check		
	Target/ Input/ Output/ Organized storage	[✓]	vSRT method:	Several vSRT methods were tested.
		[ ]	MAR system:	
		[✓]	Trial sets and physical objects:	1) Elements of datasets were defined. 2) The format of dataset was defined. See "Trial set" table.
		[✓]	Benchmarking instruments:	Benchmark indicators were designed.
		[✓]	Benchmarking results:	Technical reports were published in academic conferences.
		[✓]	Benchmarking surveys:	Several workshops were organized.
		[✓]	Benchmarking repository:	Datasets are distributed on TrackMark web.
		[ ]	External repositories:	
		Check		
Indicator	Reliability	[ ]	3DEVO:	
		[✓]	PEVO:	PEVO was defined in this activity.
		[✓]	Re-projection error of image features:	2D distance-based error of interest points was defined.
		[✓]	Position and posture errors of a camera:	1) Position error was defined using a distance between points in 3D space. 2) Posture error was defined using a degree generated with two axes.
		[ ]	Completeness of a trial:	
	Temporality	[✓]	Throughput:	
		[✓]	Latency:	
		[ ]	Time for trial completion:	
	Variety	[✓]	Number of datasets/trials:	The number of datasets was used.
		[✓]	Variety on properties of datasets/trials:	The number of variety was introduced to prevent a fine tuning for several specific datasets.

Table B.1 *(continued)*

			Check	Item	Remarks
Trial set	Dataset	Contents	[✓]	Image sequences:	Sets of still images (JPG, PNG)
			[✓]	Intrinsic/extrinsic camera parameters:	Ground truth/Reference data
			[ ]	Challenge points:	
			[✓]	Optional contents:	3D model data (used to CG images of dataset), sensor data (inertial sensors), etc.
		Metadata	[✓]	Scenario:	Film Studio, Outdoor Navigation (Campus), Indoor Navigation (Conference Venue), Work Support (Japanese Restaurant, Nursing Home)
			[✓]	Camera motion type:	Significant moving occluders Fast camera movement
			[✓]	Camera configuration:	Auto luminance control, auto focus control, etc.
			[✓]	Image quality:	1920×1080@24Hz, 720×480@Hz, etc.
	Physical object	Contents	[ ]	Physical objects:	
		Metadata	[ ]	How to find the physical objects:	

Table B.2 — Conformance checklist for evaluating template-based tracking algorithms in A.2

			Check	Item	Remarks
Process flow	Process		[ ]	Develop vSRT methods and/or MAR systems:	
			[ ]	Gather vSRT methods and/or MAR systems:	
			[✓]	Prepare and conduct benchmarking:	Metaio
			[ ]	Provide and maintain benchmarking instruments:	
			[ ]	Provide and maintain benchmarking repositories:	
			[ ]	Share benchmarking results:	
			Check		
	Target/ Input/ Output/ Organized storage		[✓]	vSRT method:	Template-based tracking algorithms
			[ ]	MAR system:	
			[✓]	Trial sets and physical objects:	See "Trial set" table.
			[ ]	Benchmarking instruments:	
			[✓]	Benchmarking results:	Providers of tracking algorithms could obtain their own benchmarking results (not shared in public).
			[ ]	Benchmarking surveys:	
			[✓]	Benchmarking repository:	It no longer exists.
			[ ]	External repositories:	
			Check		
Indicator	Reliability		[ ]	3DEVO:	
			[ ]	PEVO:	
			[✓]	Re-projection error of image features:	The lowest RMS of the re-projection error of the fiducials
			[ ]	Position and posture errors of a camera:	
			[✓]	Completeness of a trial:	Ratio of successfully tracked images
	Temporality		[ ]	Throughput:	
			[ ]	Latency:	
			[ ]	Time for trial completion:	
	Variety		[ ]	Number of datasets/trials:	
			[✓]	Variety on properties of datasets/trials:	Printed papers with various textureness levels
			Check		
Trial set	Dataset	Contents	[✓]	Image sequences:	
			[✓]	Intrinsic/extrinsic camera parameters:	Accurate ground truth poses
			[ ]	Challenge points:	
			[✓]	Optional contents:	Templates for printed papers
		Metadata	[ ]	Scenario:	
			[✓]	Camera motion type:	The camera was controlled by a robotic measurement arm.
			[ ]	Camera configuration:	
			[✓]	Image quality:	640×480@40Hz
	Physical object instances	Contents	[ ]	Physical objects:	
		Metadata	[ ]	How to find the physical objects:	

Table B.3 — Conformance checklist for benchmarking activities in A.4

		Check	Item	Remarks
Process flow	Process	[✓]	Develop vSRT methods and/or MAR systems:	HITLAB NZ
		[ ]	Gather vSRT methods and/or MAR systems:	
		[ ]	Prepare and conduct benchmarking:	
		[✓]	Provide and maintain benchmarking instruments:	HITLAB NZ
		[ ]	Provide and maintain benchmarking repositories:	
		[ ]	Share benchmarking results:	
		Check		
	Target/ Input/ Output/ Organized storage	[✓]	vSRT method:	As a sample
		[✓]	MAR system:	As a sample
		[✓]	Trial sets and physical objects:	See "Trial set" table.
		[✓]	Benchmarking instruments:	Mobile Augmented Reality Visualisation and Evaluation Toolkit for creating dataset and for make benchmarking results
		[ ]	Benchmarking results:	
		[ ]	Benchmarking surveys:	
		[✓]	Benchmarking repository:	On GitHub; 1) Open source of Mobile Augmented Reality Visualisation and Evaluation Toolkit 2) Datasets
		[ ]	External repositories:	
		Check		
Indicator	Reliability	[ ]	3DEVO:	
		[ ]	PEVO:	
		[✓]	Re-projection error of image features:	
		[ ]	Position and posture errors of a camera:	
		[ ]	Completeness of a trial:	
	Temporality	[ ]	Throughput:	
		[ ]	Latency:	
		[ ]	Time for trial completion:	
	Variety	[ ]	Number of datasets/trials:	
		[ ]	Variety on properties of datasets/trials:	

Table B.3 *(continued)*

<table>
<tr><th colspan="3"></th><th>Check</th><th>Item</th><th>Remarks</th></tr>
<tr><td rowspan="12">Trial set</td><td rowspan="10">Dataset</td><td rowspan="4">Contents</td><td>[✓]</td><td>Image sequences:</td><td>Sets of still images (80 % quality JPEG)</td></tr>
<tr><td>[ ]</td><td>Intrinsic/extrinsic camera parameters:</td><td></td></tr>
<tr><td>[ ]</td><td>Challenge points:</td><td></td></tr>
<tr><td>[✓]</td><td>Optional contents:</td><td>1) Tracking points
2) WGS84 latitude, longitude, altitude and bearing
3) Gravity, linear acceleration rotation rate and magnetic flux</td></tr>
<tr><td rowspan="4">Metadata</td><td>[ ]</td><td>Scenario:</td><td></td></tr>
<tr><td>[ ]</td><td>Camera motion type:</td><td></td></tr>
<tr><td>[ ]</td><td>Camera configuration:</td><td></td></tr>
<tr><td>[ ]</td><td>Image quality:</td><td></td></tr>
<tr><td rowspan="2">Physical object instances</td><td>Contents</td><td>[ ]</td><td>Physical objects:</td><td></td></tr>
<tr><td>Metadata</td><td>[ ]</td><td>How to find the physical objects:</td><td></td></tr>
</table>

Table B.4 — Conformance checklist for benchmarking activities in A.5

		Check	Item	Remarks
Process flow	Process	[ ]	Develop vSRT methods and/or MAR systems:	
		[ ]	Gather vSRT methods and/or MAR systems:	
		[ ]	Prepare and conduct benchmarking:	
		[✓]	Provide and maintain benchmarking instruments:	TU Graz, TUM, and UCSB
		[✓]	Provide and maintain benchmarking repositories:	TU Graz, TUM, and UCSB
		[ ]	Share benchmarking results:	
		Check		
	Target/ Input/ Output/ Organized storage	[ ]	vSRT method:	
		[ ]	MAR system:	
		[✓]	Trial sets and physical objects:	See "Trial set" table.
		[ ]	Benchmarking instruments:	
		[ ]	Benchmarking results:	
		[ ]	Benchmarking surveys:	
		[✓]	Benchmarking repository:	3D model data for paper craft buildings, paper folding plans, video sequences, etc. are distributied on TU Graz website.
		[ ]	External repositories:	
		Check		
Indicator	Reliability	[ ]	3DEVO:	
		[ ]	PEVO:	
		[ ]	Re-projection error of image features:	
		[ ]	Position and posture errors of a camera:	
		[ ]	Completeness of a trial:	
	Temporality	[ ]	Throughput:	
		[ ]	Latency:	
		[ ]	Time for trial completion:	
	Variety	[ ]	Number of datasets/trials:	
		[ ]	Variety on properties of datasets/trials:	
		Check		

Table B.4 *(continued)*

			Check	Item	Remarks
Trial set	Dataset	Contents	[✓]	Image sequences:	Video sequences (avi)
			[✓]	Intrinsic/extrinsic camera parameters:	Ground truth acquired using a robot arm
			[ ]	Challenge points:	
			[✓]	Optional contents:	3D model data for paper craft buildings
		Metadata	[ ]	Scenario:	
			[✓]	Camera motion type:	1) Robotic arm motion or free hand-held motion 2) Birds view, top view, and street view
			[ ]	Camera configuration:	
			[✓]	Image quality:	1600×1200 or 640×480 Several lighting conditions
	Physical object instances	Contents	[✓]	Physical objects:	The following paper craft buildings used as physical objects are as follows: 1) The Pyramid of Cheops (also called the Pyramid of Khufu), 2) The Berliner Dom (Berlin Cathedral), 3) The Arc de Triomphe de l'Etoile in Paris, 4) The Musikverein in Vienna (Vienna concert hall), 5) A medieval Irish Round Tower, 6) St. Mark's Campanile in Venice.
		Metadata	[✓]	How to find the physical objects:	Paper folding plans

Table B.5 — Conformance checklist for scenario 1 in A.6

		Check	Item	Remarks
Process flow	Process	[✓]	Develop vSRT methods and/or MAR systems:	Competitors
		[ ]	Gather vSRT methods and/or MAR systems:	
		[✓]	Prepare and conduct benchmarking:	Competition organizers
		[ ]	Provide and maintain benchmarking instruments:	
		[ ]	Provide and maintain benchmarking repositories:	
		[✓]	Share benchmarking results:	Competition organizers
		Check		
	Target/ Input/ Output/ Organized storage	[✓]	vSRT method:	
		[✓]	MAR system:	
		[✓]	Trial sets and physical objects:	See "Trial set" table.
		[ ]	Benchmarking instruments:	
		[✓]	Benchmarking results:	Results are announced onsite.
		[ ]	Benchmarking surveys:	
		[ ]	Benchmarking repository:	
		[ ]	External repositories:	
		Check		
Indicator	Reliability	[ ]	3DEVO:	
		[✓]	PEVO:	1) An element defined by 3D coordinates is correctly identified or not. 2) Overlay quality is checked by the organizer.
		[ ]	Re-projection error of image features:	
		[ ]	Position and posture errors of a camera:	
		[ ]	Completeness of a trial:	
	Temporality	[ ]	Throughput:	
		[ ]	Latency:	
		[✓]	Time for trial completion:	
	Variety	[ ]	Number of datasets/trials:	
		[✓]	Variety on properties of datasets/trials:	Speed and variance of the rotation

Table B.5 *(continued)*

			Check	Item	Remarks
Trial set	Dataset	Contents	[ ]	Image sequences:	
			[ ]	Intrinsic/extrinsic camera parameters:	
			[✓]	Challenge points:	Challenge points on the base underneath the vehicle model
			[✓]	Optional contents:	1) 3D coordinates of challenge-point candidates 2) 3D coordinates of reference points
		Metadata	[✓]	Scenario:	Tracking of a rotating vehicle model
			[✓]	Camera motion type:	Variable rotation speed (fixed camera and rotated vehicle model)
			[ ]	Camera configuration:	
			[ ]	Image quality:	
	Physical object instances	Contents	[✓]	Physical objects:	1) 1:10 vehicle model with a rotating vehicle base 2) Reference points used only for calibrating (not available in the competition phase)
		Metadata	[ ]	How to find the physical objects:	