

New Work Item Proposal: A Standard Reference Model for Generic MAR Systems



ISO JTC 1 SC 24 WG9

**GERARD J. KIM
KOREA UNIVERSITY**

What is a Reference Model?



- A reference model (for a given domain) defines an authoritative basis that outlines:
 - Set of principles
 - Terms and their precise definitions
 - Generic system model of mixed/augmented reality system
 - ✦ Major components and their functionalities
 - ✦ Inter-component interfaces (data and control)
 - ✦ @ the right abstraction level w.r.t. purpose
 - Content model and file format for MAR
 - ISO SC24 WG9 (“Augmented Reality Continuum (ARC)”)
 - Validation use cases
 - Extensions: “Reference modules”

Principles (1)



- AR/MR is “implemented” as **VR** system
 - Relevance of SC24
- Level at par with other standards
 - Use existing standards (e.g. data compression)
 - For harmonious future integration (W3C/HTML, Web3D/X3D, etc.)
- User view: Content developer > Implementer
- Be able to describe a reasonable range of “ARC” applications

Principles (2)



- Independence from *specific implementations*
 - Algorithms: E.g. Recognition/tracking, Rendering, ...
 - Sensors: E.g. Camera vs. RFID, ...
 - Platform / Distribution of computation: E.g. Desktop, Server-client, Cloud, ...
 - Real world capture: E.g. Camera vs. Kinect
 - ✦ 2D Video as abstraction of the “Real World”? (what about 3D video?)
 - Virtual/Mixed reality world
 - ✦ Abstract scene graph
 - Output Displays: E.g. HMD, Mobile, Projector, Holography, ...
 - ✦ Abstracted as parameterized image plane (projection of a “scene”)
 - ✦ Extensions in the dimension of modality: Visual, Aural, Haptic, ...

Terms (in the context of MAR)

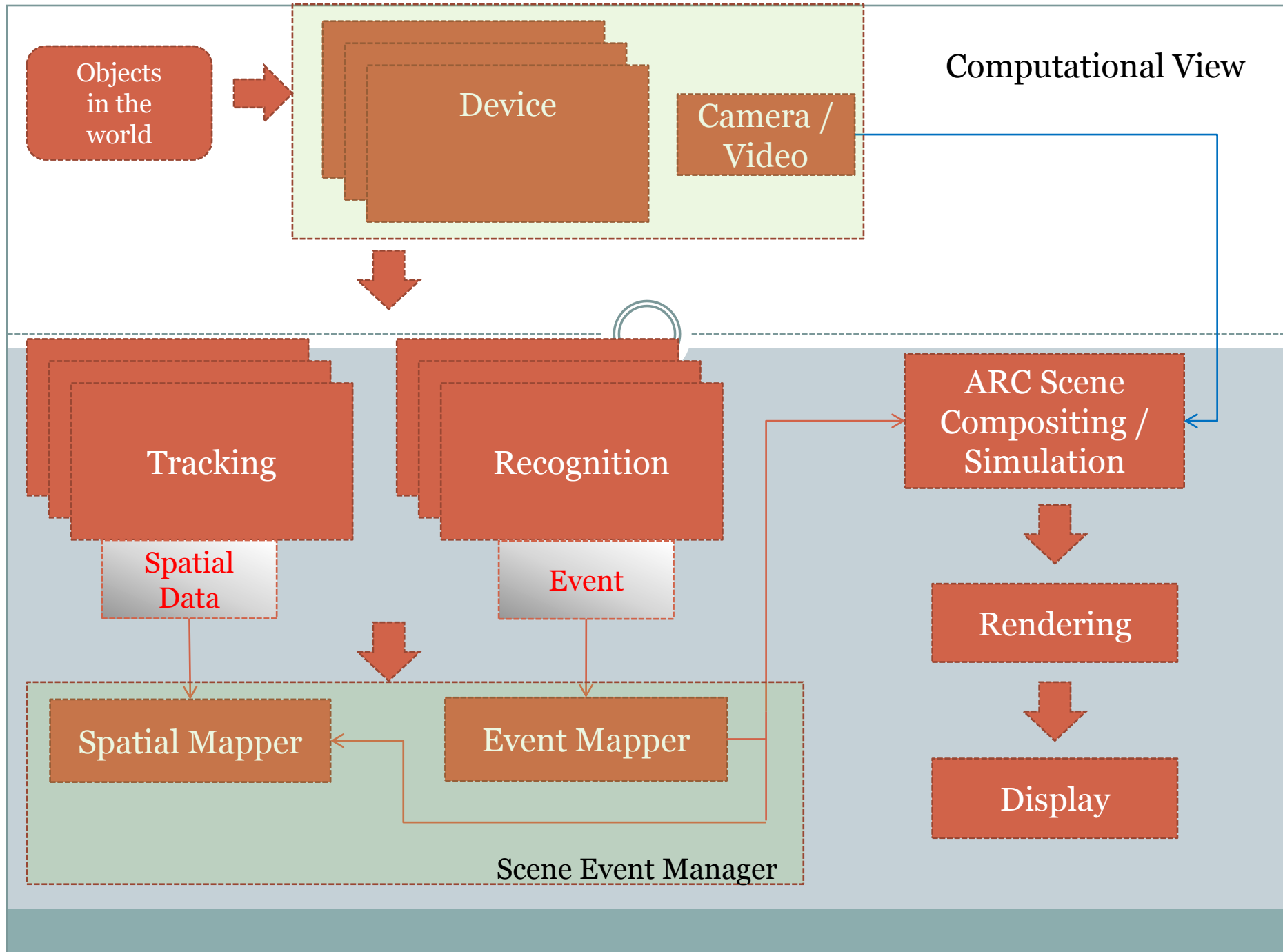


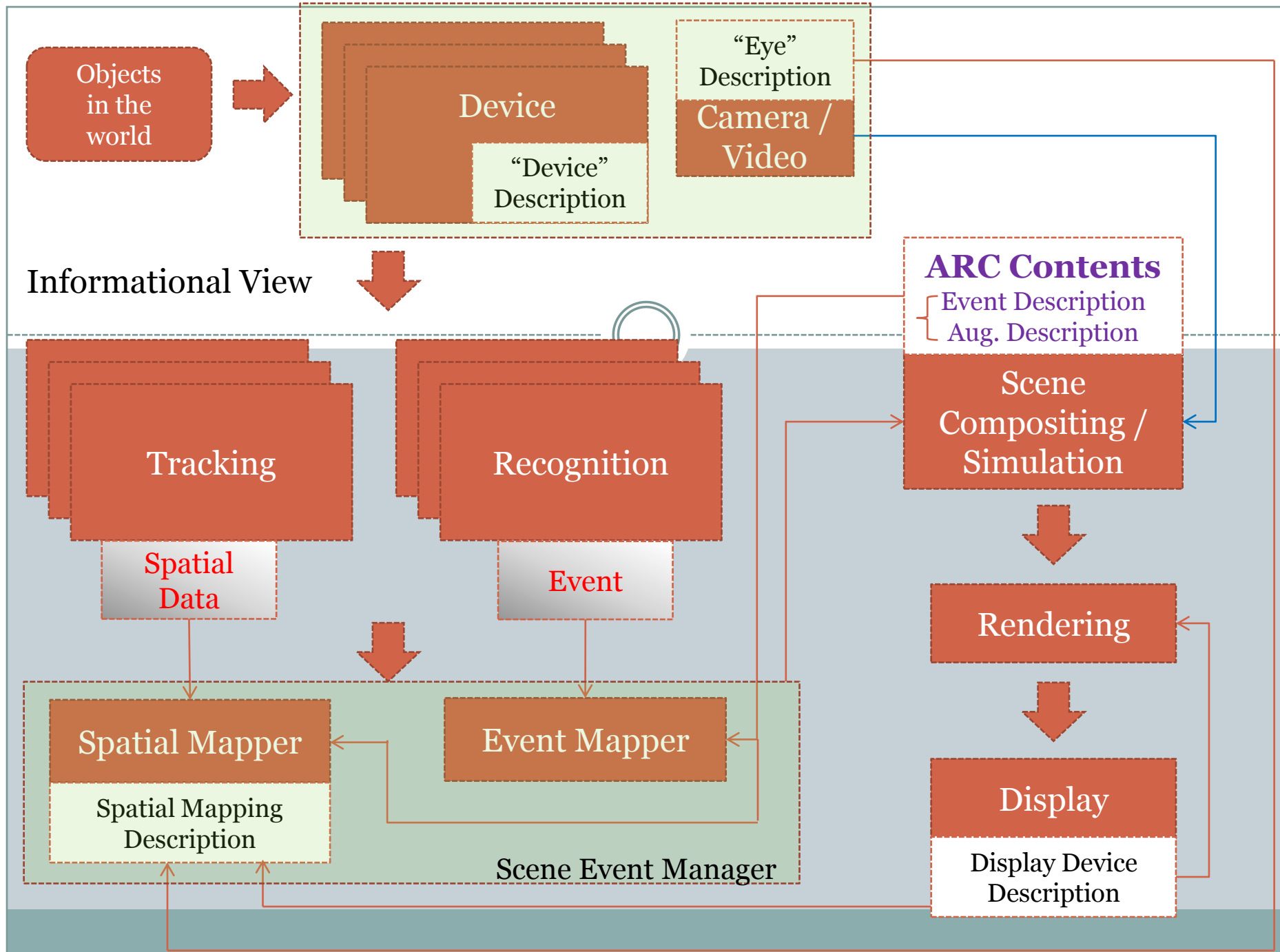
- Augmented/Mixed/Virtual Reality
- Mixed (Augmented) Reality Continuum
- Sensors, Devices, Tracking, Recognition, ...
- Event/Context, ...
- Augmentation, Behavior, Resources, Elements, ...
- Scene, Display, “Eye”, ...
- Real world capture
- Mapping
- AR/MR Content, Association
- Browser, Player, Application, ...

Proposal



- Browser chooses the algorithms
 - Tracking
 - Rendering
 - Display adaptation
- **AR Contents = A Set of {Events, Augmentation} Associations**
 - Events = Context, conditions, ...
 - Augmentation = VR objects, 2D text, animation, behaviors, ...
 - ✦ Spatial information = How to spatially register augmentation in real space
- Need a protocol to define standard “events” and their mappings between the browser and content
- Sensors - Optional





Sensors (Devices)



- A sensor is a hardware "device" that measures a physical quantity and converts it into a "raw" signal which can be read by another module
- Sensor ("device") description - a declarative description that describes the type of the sensor device, its important attributes and values
- Attributes of a sensor can include,
 - Sensor abstract category (e.g. imaging, gps, rfid, depth, ...)
 - Important parameters of the sensor (e.g. focal length, sampling rate, ...)
 - Aspect of the target physical world or object the sensor intended to measure (e.g. position, depth, orientation, ...)
- Input: No direct input (Real world itself as it is ..)
- Output: The raw signal
 - Depends on the type of the sensor used (e.g. binary image, color image, depth map, ...)

Recognition Module



- A software module that takes raw sensor device data and produces "events" that match the description given by the content specification with the same identifier.
- The event description must be described in a standard protocol, language, and naming convention. E.g. The content specification might define an event as:

Identifier	"Event 1", "Location 1", "My_Event", ...
Type	Location, Object, Marker, Face, ...
Value	(100, 100), Apple, HIRO, John_Smith, ...
...	...

- Input
 - Raw sensor device data
 - Event Description
- Output: Event data

Tracking Module



- A software/hardware module that takes raw sensor device data and produces the position and orientation of the target physical object or entity which is designated by the event description from the content specification.
- The event description must be described in a standard protocol, language, and naming convention. E.g. The content specification might define an event as:

Identifier	“Event 1”, “Location 1”, “My_Event”, ...
Type	Location, Object, Marker, Face, ...
Value	(100, 100), Apple, HIRO, John_Smith, ...
Tracking data	Inertial position, 4x4 Transformation matrix, ...

- Input
 - Raw sensor device data
 - Event Description
- Output: Spatial data (in different formats)

Scene Event Manager



- A software module that takes external events and simulates the scene behavior and dynamically updates the AR scene description accordingly
- The behavior of the dynamic AR scene is specified in the content description
- Collectively composed with the Event Mapper and Spatial Mapper

Event Mapper



- A software module that relays the event produced by the "Recognition" module to the "AR scene event manager"
- It also parses the event description and lets the recognition module understand which event to be recognized for the specified content.
- The events are defined in the given content specification
- Input
 - Events from the Recognition module
 - Event description
- Output
 - Event invocation call to the scene data
 - Event definition call to the Recognition module

Spatial Mapper



- A software module that relays the tracking data produced by the "Tracking" module to the "AR scene event manager"
- It also parses the event description and lets the tracking module understand which event to be recognized and which object to be tracked for the specified content
- It also takes the External Camera/Video description and maps its specification into the virtual camera into the scene
- **Input**
 - Tracking data from the tracking module
 - Event description
 - Camera description
- **Output**
 - Event/tracking update call to the scene data
 - Tracking event definition call to the Tracking module
 - Camera position setting call to the scene data

MAR Content Scene / Execution Platform



- A dynamic hierarchical data structure that describes the virtual scene. For AR purpose, the content scene is the traditional scene graph for virtual world added with declarations for:
 - AR events, AR sensor device, AR camera, AR display capabilities
- The MAR Content (in the scene) can be specified using:
 - X3D/HTML5/MPEG4 + new constructs for above
 - **Completely new constructs**
- The execution platform "example" may be:
 - Basic scene graph renderer
 - + Additional AR functionalities (mapping) implemented by DOM
 - + Other browser specific implementations
- **Input**
 - External events (can include other usual device events such as mouse input)
- **Output**
 - Updated Scene Graph

Camera/Eye Module



- Special type of sensor
- A real world capturing device is a hardware/software that produces a video stream (and other “visual” data format) to be embedded into the AR scene.
 - Camera
 - Video streamer
 - Static image background
 - Real world (e.g. Optical see through case)
- Camera/Eye description attributes and values of the virtual eye for the real world visual data
 - FOV
 - External/Internal parameters
 - Resolution
 - Parent coordinate system
- Input: None
- Output: Video stream

Renderer

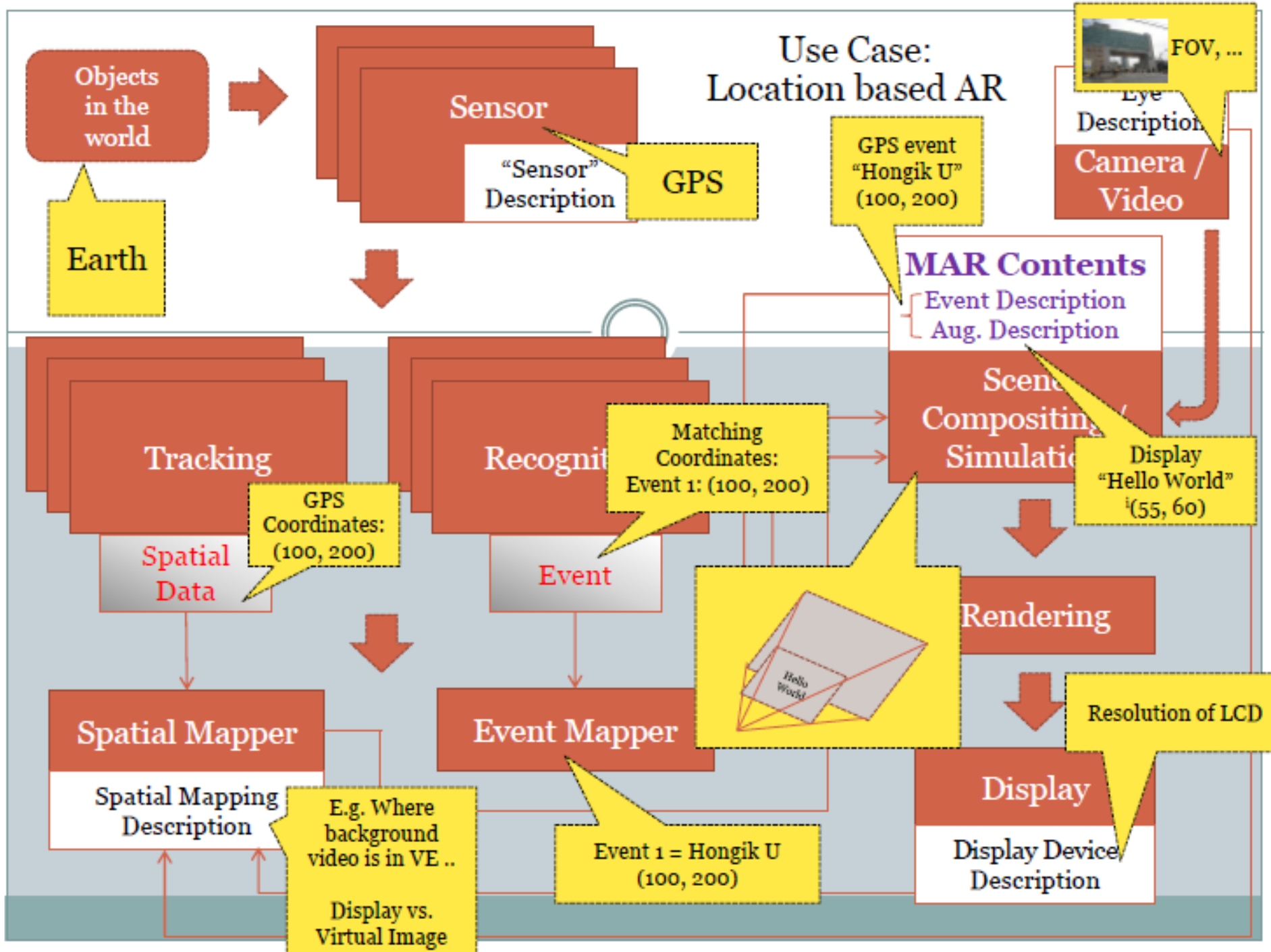


- Renderer takes the scene graph and produces rendering signal multimodally (visual, aural and haptic)
- It renders according to the display device description
- Input
 - Scene description
 - Display description
- Output
 - Rendering signal

Display



- A hardware device that displays the scene in different modalities (visual, aural, and haptic)
- It is associated with description outlining its type and important parameter and values regarding its capabilities
 - Visual: size, resolution, color space, ...
 - Audio: amplitude range, frequency range, ...
 - Haptic: sampling rate, force output range, operating range, ...
- Input
 - Rendering signal
- Output
 - Displayed contents



AR/MR Content Model



- **Context**
 - Conditions for with augmentation to occur
 - “AR Events”
 - ✦ Marker recognition
 - ✦ Location recognition
 - ✦ ...
- **Augmentation**
 - 2D → HTML?
 - 3D → X3D?
 - Other: Haptic, Sound, ...
- **Context + Augmentation**
 - New constructs
 - X3D nodes
 - HTML elements?

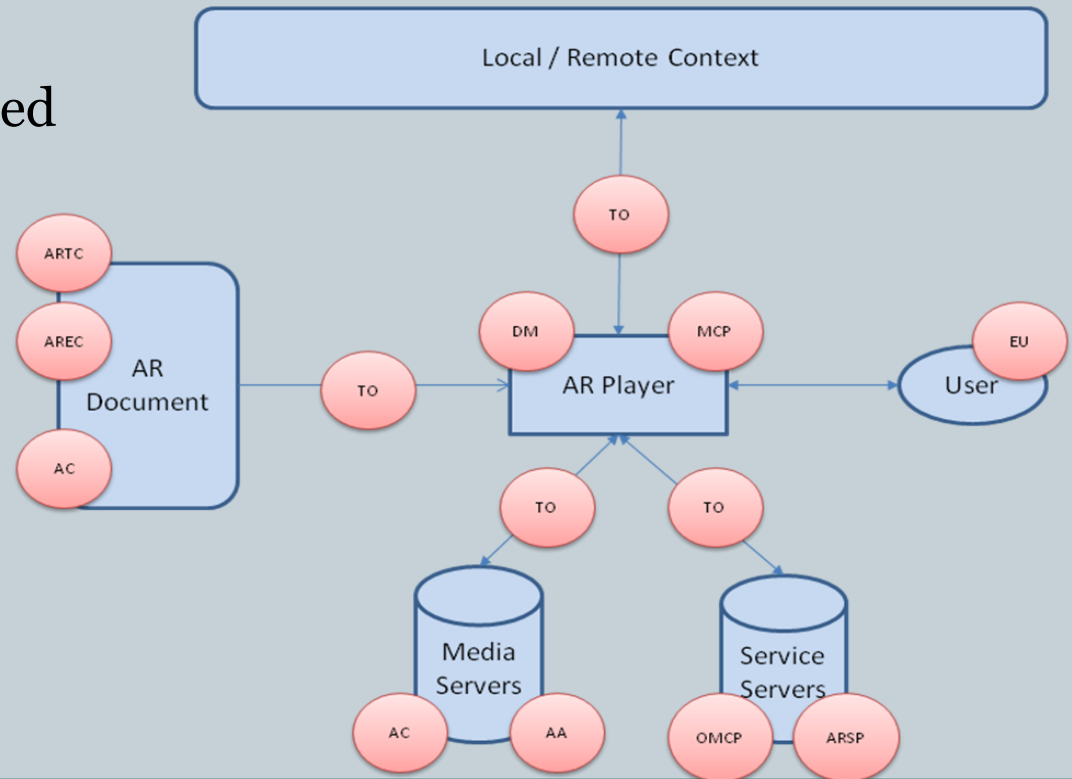
Reference Modules



- **Refinement of the functional modules**
 - Clarify its purpose and functionalities at a lower level
 - Maintain generality
 - Address applicability
 - Relationship with other modules at the lower level
 - Development into an “Application” Reference Model
- **E.g.**
 - Physical sensor (Device) module
 - Modules refinement for video avatars and interaction
 - Script engine (Mapper) module

Making the RM More Complete ...

- Views – George Percivall (OGC)
 - ISO Approach to Reference Models
 - 5 Views: Computational, Informational, Enterprise, Engineering, Technology
- SC29/ARS Work to be merged
 - More abstract level RM
 - Other views
 - ✦ Enterprise view
 - ✦ Engineering view
 - Other modules
 - ✦ E.g. Media server
 - More use cases



Making the RM More Complete ...



- **Asset DB**
- **Calibration between virtual and physical worlds / Units**
- **Displays**
 - Projectors (and projective textures)
 - See through HMD (e.g. Google glass)
- **Performance benchmarking**
 - What to test for (modules and performance criteria)
 - How to test (procedure)
 - Adherence
 - ✦ In addition to structure and functionalities
- **Applications of RM**
 - File formats (information) → WG6, Web3D, W3C, SC29, ...
 - Reference modules
 - Implementations / Use cases

Integration with Existing Standards / Collaboration with other SDO's



- **X3D / Web3D (SC24)**

- Already has a rich and mature 2D/3D representation scheme and file format
- Can be used as scene representation for AR (which is really VR space)
- Can be used for 2D/3D object representation and their behaviors (X3DOM, Behavior nodes, etc.)
- Working closely with Web3D AR WG

- **SC29 / ARS**

- Pursuing its own AR RM

- ✦ Based on work by ARS

- Ad-hoc standards group lead by Perey Research Associates
- Has many industrial sentimental? grass-roots type of support

- ✦ Talks are on-going to merge the work and specialize in respective areas of expertise (subject to approval by SC24)

- SC29 Expertise: Recently highlighted mission regarding AR

- ✦ Online and real time support (e.g. compression and streaming)
- ✦ Multi-sensorial experiences (e.g., haptics and olfactory) ?
- ✦ Extended audio-visual experiences (e.g., 3D video and 3D audio)
- ✦ 3D scene representation ?

Integration with Existing Standards / Collaboration with other SDO's



- W3C / HTML 5
 - POI
 - WebGL / Declarative 3D
 - **Trend: Web is “housing” everything**
 - ✦ Video, Audio, 3D Virtual, Documents, Interactivity, ...
 - ✦ Web browser vs. MPEG browser vs. X3D browser
- Multi-SDO Standardization Effort
 - Put forth by Neil Trevett (KHRONOS)

Roadmap



- Resolve issues with SC29
 - Co-publishing of the RM (during this meeting)
- Work items for SC24: 3D augmentation contents
 - AR reference model (with focus on “SC24” areas)
 - Reference modules
 - ✦ Device model
 - ✦ Modules for video avatar and interaction
 - AR benchmarking
- Continued Refinement of the Ref. Model
 - More use cases and implementations
 - Documentation
 - Merging with SC29
 - CD by October (?)
- Information constructs/ File format proposal (based on the RM)
 - Through WG6
 - AR Events, Devices, AR Avatar, ...