Rich Media Standards in MPEG

Convenience by Convergence

Young-Kwon LIM
young@netntv.co.kr
Rich Media?

- mono media
- multi media
  - video + audio
- rich media
  - (animating) image
  - (vector) graphics
  - interaction
- (future) media

Convenience by Convergence
Contents

• MPEG-4 BIFS (ISO/IEC 14496-1)
• MPEG-4 LASeR (ISO/IEC 14496-20)
BIFS (Binary Format for Scene)

- A scene description for composition is needed because MPEG-4 is an object-based representation system.

- BIFS is the MPEG-4 scene description protocol:
  - to compose MPEG-4 objects (temporal & spatial)
  - to describe interaction with MPEG-4 objects
  - to animate MPEG-4 objects
Components of BIFS

DELIVERY

DECODING

SCENE GRAPH MANAGEMENT

PRESENTATION

BIFS-Update ES

BIFS Anim ES

MPEG-4 Streams

VRML Nodes

MPEG-4 Nodes

2D Nodes

Audio Nodes

3D Nodes

2D+3D Nodes

Interaction

S&N Sound

FBA

Rendering

Convenience by Convergence
Coordinate System
Basic Primitives

- Rectangle
- Circle
- IndexedFaceSet2D
- IndexedLineSet2D
- PointSet2D
- Curve2D
- Bitmap
- Text
Routes and Events

- Events are usually generated (eventOut) by sensor nodes (click, drag, periodic events, ..) and shall be connected to event listener (eventIn) in order to modify the scene: this connection is called a Route.
- A field can generate events (eventOut), accept event as an input (eventIn), do both (exposedField) or none (field).

<table>
<thead>
<tr>
<th>BIFS Node</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TouchSensor</td>
<td>Detects mouse move and button actions and generates position events, time events and state events (over, clicked)</td>
</tr>
<tr>
<td>DiscSensor</td>
<td>Detects mouse move and translates them into rotation events. The rotation center is the origin of the local coordinate system.</td>
</tr>
<tr>
<td>PlaneSensor2D</td>
<td>Detects mouse move and translates them into relative translation events</td>
</tr>
<tr>
<td>InputSensor</td>
<td>Detects predefined events outside the scene and execute a list of BIFS commands upon reception. The events can be sent by a keyboard, a mouse, a joystick, etc.</td>
</tr>
<tr>
<td>ProximitySensor2D</td>
<td>Detect mouse presence in a virtual rectangle and generates position events, time events and state events (is inside)</td>
</tr>
</tbody>
</table>
BIFS Update

• Concept of BIFS Update
  – BIFS scenes may be modified through commands.
  – Commands are grouped in a timed entity called Access Unit in MPEG-4 Systems
  – A BIFS scene is made of a sequence of BIFS-Commands access units
  – Playing a BIFS stream consists in applying modification described in the BIFS commands to the scene and displaying the result to the user.

• BIFS-Anim
  – BIFS Command frames, consists in modifying node fields or adding or suppressing parts of the scene graph.
  – BIFS Animation frames, consists in modifying a pre-defined set of node fields with high compression techniques, and is designed for continuous animation.
BIFS service in Korean T-DMB
BIFS service in French Digital Radio
• ISO/IEC 14496-20

• **Lightweight Applications Scene Representation**
  – Scene Description for a lightweight 2D rich media presentation
  – Simple Aggregation Format for logical multiplexing
Target Applications

• mobile interactive streamable rich-media services
  – Mobile broadcasting with integrated interactive data
  – Cartoon with efficient vector graphics
  – Advanced navigation with incremental contents

• Specific requirements for handheld devices
  – Low bitrate content: Rich Media from 10 Kb/s
  – Low footprints implementation from 30 Kb Java, <100 Kb C++
  – (Streamezzo showed implementation with 75 Kb)
SVG (Scalable Vector Graphics)

- SVG is a language for describing two-dimensional graphics in XML.
  - Three types of graphical objects: vector graphic shapes, images, and text
  - Grouped, styled, transformed and composited into previously rendered objects.
  - Nested transformations, clipping paths, alpha masks, filter effects, template objects and extensibility.

- Three major improvements to Web
  - Data-driven graphics
  - Interactive graphics
  - Personalized graphics

- SVG Mobile (Tiny)
  - Full > Basic > Mobile (tiny)
  - Mobile devices having different characteristics in terms of CPU speed, memory size, and color support
  - Constraints on content, attribute types, properties, and user agent behavior.
LASeR vs SVG

- Extended Scene Description
  - Integration of audio and video from SMIL
  - Inclusion of XML Event listener
  - Restriction & extension to existing element
    - animation elements are not animatable and inheritable.
    - g elements for selection, clipping, simple matrix layout
    - Support of underlined text
  - Support of OpenType font (ISO/IEC 14496-18)

- Synchronized dynamic management of scene
  - Commands with specific time stamp attached
  - Update commands: Add, Insert, Delete, Replace
  - Intra/Inter: NewScene, RefreshScene
Enhanced Scene Description (1/2)

- Scene segment for "conditional" timeline & personalized contents
  - When a scene segment starts with a NewScene, the scene time is reset to 0.
  - When a scene segment does not start with a NewScene, the scene time is not reset to 0 and let \( T_0 \) be the scene time within the initial scene segment upon reception of the first access unit of that new scene segment.

\[
\text{sceneTime}(x) = \text{mediaTime}(x) - \text{mediaTime}(\text{NewScene})
\]

\[
\text{sceneTime}(y) = \text{mediaTime}(y) - \text{mediaTime}(\text{NewScene})
\]
Enhanced scene description (2/2)

• **Enhanced support of text scrolling**
  – scroll a set of objects of unknown size, inside a clipping rectangle.
  – define the speed of the scrolling in terms of clip size, not in terms of the objects size
  – define the bounds of scrolling precisely, according to the actual size of the objects
  – define the beginning and ending conditions (screen full or empty of the objects)
  – define automatic and manual scrolling (continuous scrolling, page advance on action...), and allow a combination of automatic and manual
  – define scroll stops (for manual scrolling) by page (size of viewport) or by object markers

• **Conditional**
  – set of commands inserted by the user events or time events
  – simple logics without use of scripting for easy authoring

• **Full screen video**
  – specific attribute for full screen rendering of video element for the efficient use of hardware support
Streaming & Broadcasting support (1/2)

• Time-based differential encoding of scene description
  – Differential encoding of scene description similar to video and audio encoding
  – LASeR Commands carrying scene description at the specific time instance
  – NewScene command for carrying initial scene (no dependencies)
  – Update commands carrying differences between the previous scene and the current scene
    • Add/Insert/Delete/Replace
    • Save/Restore/Clean
    • Activate/Deactivate
  – RefreshScene command containing the current status of the scene for independent decoding

• Binarization of scene description
  – Efficient delivery (better than GZIP)
  – Fast processing at the decoder (no parsing of “string”)
Streaming & Broadcasting support (2/2)

- StreamSource to improve channel switching behaviour
  - Usage of mosaic channel to minimize the time for black screen
  - Alternate contents for the period of channel switching

- Referencing resources across the presentation
  - GlobalIDs are defined for the unique referencing
Improved user interface support for handheld devices

• **CursorManager**
  – Virtual pointer emulating the mouse for the phones without pointing devices
  – Defining the movements of mouse pointer when the specific keys are selected

• **Screen orientation**
  – For handheld devices for rotatable screens

• **Animation with external sources**
  – Animation synchronized with external sources
  – Remaining batteries, reception qualities, and etc...

• **Abstract events**
  – Definition of "events" not linked to specific "keys"
  – For the support of different mapping of keys between the products or vendors
Functional requirements of SAF

- Tight synchronization with audio or video contents for the services associated to audio or video contents
- Simple aggregation mechanism for the various media assets comprising rich media contents
- Signaling of life time of resources for the reuse of media clips and optimization of device resources
Basic structure of SAF packet

- Fixed length header for easy processing
- `accessUnitLength` : self-contained packet structure for simple delivery & storage
- `compositionTimeStamp` : tight synchronization with audio or video
- `fragmented AU` : simple aggregation of various media assets

<table>
<thead>
<tr>
<th>1</th>
<th>15</th>
<th>1</th>
<th>1</th>
<th>30</th>
<th>16</th>
<th>4</th>
<th>12</th>
<th>(accessUnitLength – 2) * 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>randomAccessPointFlag</td>
<td>AU_sequenceNumber</td>
<td>presentOPTS</td>
<td>presenceEXTS</td>
<td>compositionTimeStamp</td>
<td>accessUnitLength</td>
<td>accessUnitType</td>
<td>StreamID</td>
<td>payload</td>
</tr>
</tbody>
</table>

Convenience by Convergence
Stream Management

- Aggregation of packets and streams
  - accessUnitType to distinguish different types of packets
  - streamID to distinguish different sources
  - No needs for out-of-band signaling

- Signaling of “end” for resource optimization
  - EndOfStream
  - EndOfSAFSession

<table>
<thead>
<tr>
<th>Type of access unit payload</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x00</td>
<td>Reserved</td>
</tr>
<tr>
<td>0x01</td>
<td>TransientStreamHeader</td>
</tr>
<tr>
<td>0x02</td>
<td>NonTransientStreamHeader</td>
</tr>
<tr>
<td>0x03</td>
<td>EndOfStream</td>
</tr>
<tr>
<td>0x04</td>
<td>AccessUnit</td>
</tr>
<tr>
<td>0x05</td>
<td>EndOfSAFSession</td>
</tr>
<tr>
<td>0x06</td>
<td>CacheUnit</td>
</tr>
<tr>
<td>0x07</td>
<td>RemoteStreamHeader</td>
</tr>
<tr>
<td>0x08</td>
<td>GroupDescriptor</td>
</tr>
<tr>
<td>0x09</td>
<td>FirstFragmentUnit</td>
</tr>
<tr>
<td>0x0A</td>
<td>FragmentUnit</td>
</tr>
</tbody>
</table>

---

Convenience by Convergence
Resource Management

- Multiple use of resource
  - TransientStreamHeader: the stream shall be used once and by only one media element.
  - Non-transientStreamHeader: the stream shall be used multiple times.

- Cache Unit
  - Delivering and storing cache objects for reuse across the presentation

- Use of external resources
  - RemoteStreamHeader to refer the external resources with URL
  - GlobalStreamID to refer the streams across the presentation
Broadcasting Support

- Random access & Carouselling
  - Random access point indication by randomAccessPointFlag
  - Carouseling of configuration packets
  - Repetition of data packet with sequenceNumber

- Fast channel switching by using multiple decoding buffers
  - Buffering multiple streams for quick start of decoding.
  - Explicit signaling of grouping of streams for buffering
LASeR in OMA and 3GPP

- OMA RME
  - Work Item on Use cases and Requirement for Rich-Media Environment
  - Supported by IBM, Orange, Three, Bouygues Telecom, Alcatel, TNA, Medialive and Streamezzo
  - Finalisation of the Requirement Document in September 2005

- 3GPP SA4 DIMS
  - Work Item to specify dynamic interactive rich-media scene
  - Supported by Apple, Three, T-mobile, Orange, Vidiator, Siemens, Streamezzo
## Comparison of Components

<table>
<thead>
<tr>
<th>Component</th>
<th>BIFS ISO/IEC 14496-1 &amp; 11</th>
<th>Component</th>
<th>LASer ISO/IEC 14496-20</th>
</tr>
</thead>
<tbody>
<tr>
<td>VRML with extensions</td>
<td>Scene Description</td>
<td>Scene Compression</td>
<td>SVGT1.2 with Extensions</td>
</tr>
<tr>
<td>BIFS commands</td>
<td>Update Mechanism</td>
<td>Update Mechanism</td>
<td>LASer commands</td>
</tr>
<tr>
<td>context-based binarization</td>
<td>Scene Compression</td>
<td>Scene Compression</td>
<td>simple binarization</td>
</tr>
<tr>
<td>SL packet</td>
<td>Packetization</td>
<td>Packetization</td>
<td>SAF packet in specially configured SL packet</td>
</tr>
<tr>
<td>OCR, DTS, CTS</td>
<td>Timing</td>
<td>Timing</td>
<td>CTS only</td>
</tr>
</tbody>
</table>
Conclusion

• Presentation technologies for convenient use of rich media
  – Integration of various media in one screen
  – Seamless use of various media
  – Fast and efficient update of look & feel

• Summary of each technologies
  – MPEG-4 BIFS is powerful and generic presentation language supporting 2D/3D
  – MPEG-4 LASeR is successor of MPEG-4 BIFS specialized for handheld devices with limited resource
Thank You!

Questions?