To Maximize Interoperability in Mobile News Gathering

Yoshiaki Shibata[1] and Nobuo Kamata
Soliton Systems K.K.
Shinjuku, Tokyo, Japan
yoshiaki.shibata@soliton.co.jp

Abstract – *This paper introduces the standardization activities in SMPTE, which maximizes interoperability among applications for the mobile news gathering. To meet a demand on industry standard basic metadata items to be obtained during acquisition, an evolution of the SMPTE UMID is discussed not only for the basic description of "When/Where/Who has created a frame" but also for the camera shooting direction. Some other SMPTE activities relevant to the mobile news gathering are also discussed.*

INTRODUCTION

A drastic evolution of public mobile network enables even a smartphone to plays a crucial role for the mobile news gathering. In addition, an appearance of a drone further accelerates the mobility at shooting. In such a situation, a use of ancillary information items such as those obtained from the GPS unit and/or a gyro sensor attached to a camera during acquisition has become a common practice and various proprietary solutions have been spreading.

Now, it's time to seriously consider an interoperability of those information items as the basic metadata to gain synergy effects among products from multiple vendors.

In this paper, the latest standardization activities in SMPTE to maximize the interoperability are introduced with an emphasis on the SMPTE UMID (Unique Material IDentifier), which is now evolving to describe not only "When/Where/Who has created each frame" but also the camera shooting direction for the frame in a media stream.

In addition, a couple of activities in SMPTE relevant to the mobile news gathering applications are also discussed.

UNIQUE MATERIAL IDENTIFIER (UMID)

The UMID is a globally unique audiovisual material identifier standardized by the SMPTE as SMPTE ST 330 [1] and RP 205 [2]. In 1998, the EBU/SMPTE Joint Task Force recommended that a unique material identifier and a generic file wrapper were needed to facilitate the exchange of program material and metadata as bit-streams [3]. These recommendations were realized by the standardizations of the UMID and the MXF (Material eXchange Format) [4] in the year 2000 and 2004, respectively, which have constituted the basis of the file-based media production workflow we have today.

Figure 1 shows the UMID structure. A UMID (called Extended UMID) is composed of a 32-byte Basic UMID as a globally unique material identifier and a 32-byte Source Pack describing "When/Where/Who has created a material,"

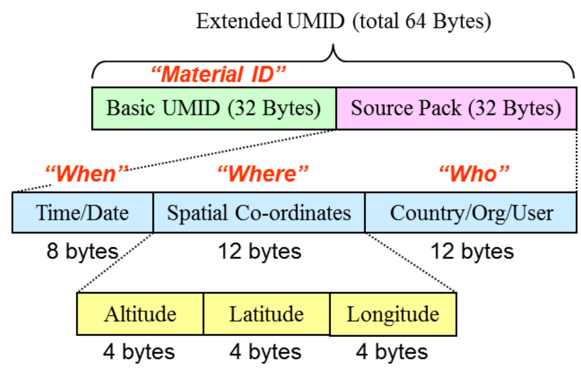


FIGURE. 1: UMID STRUCTURE.

in which the "Where" part is further decomposed into the Altitude, Latitude and Longitude components.

Figure 2 demonstrates how the UMID is basically used in the mobile news gathering. In this figure, a wireless camera equipped with the GPS unit and gyro sensors generates and wirelessly transfers a media stream as a sequence of frames with UMIDs being interleaved. The media stream is then received by a video receiver which converts the media stream into the SDI (Serial Digital Interface) output with UMIDs being attached to the respective frames based on SMPTE ST 291 [5]. An MXF video recorder is provided to receive arbitrary portion of the SDI input and to record it as an MXF file.

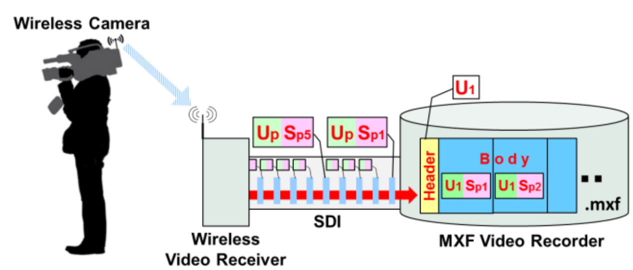


FIGURE. 2: BASIC UMID APPLICATION IN MOBILE NEWS GATHERING

As shown in the figure, each frame is attached with a UMID composed of the Basic UMID (U_p) and the Source Pack (S_{pj}, $j=1, 2, \ldots$). Note that the Basic UMID is common to all frames in a media stream in order to uniquely identify the media stream, while the Source Pack varies with frames (at its "When" and "Where" parts). As the granularity of "When" part is specified so that it is smaller than the

[1] Chair, SMPTE TC-30MR Study Group on UMID Applications.

reciprocal of the frame rate, each frame in a media stream is globally uniquely identified by the UMID (as a combination of a globally unique Basic UMID and a Source Pack unique within the media stream).

An MXF file is composed of a (file) Header containing various metadata items and a (file) Body storing a bounded sequence of (compressed) frames. To globally uniquely identify an MXF file as a material, the Basic UMID contained in the file Header is used [4]. Furthermore, an (Extended) UMID, called Body UMID, is also stored in the file Body by recording it next to a corresponding frame.

When the SDI input is received to create an MXF file as shown in Figure 2, a Basic UMID to globally uniquely identify the MXF file (U_1) is generated and stored in the file Header. This Basic UMID value is also used for the Basic UMID of the Body UMID for a frame in the file Body because the bounded sequence of frames stored in the file Body is equivalent with the MXF file by itself as a material.

The Source Pack in the Body UMID, on the other hand, inherits the one attached to an incoming frame *as is*. This is because of the Source Pack describing the origin of a material by definition or, "When/Where/Who has originally created a frame."

As a result, when the recorder receives an incoming frame attached with a UMID of "$\underline{U_p}\ S_{p1}$" in the SDI input, it stores the frame (after its compression as needed) with the Body UMID of "$\underline{U_1}\ S_{p1}$" into the file Body.

UMID EVOLUTIONS

I. UMID for HD in 2004

In the year 2004, a major update was made for the UMID format specification so that an altitude is expressed not only as a value from the center of the earth but also as a signed value relative to the sea level of the local geoid. In addition, a description of supplemental information such as the number of satellites used for the GPS measurement, a use of any supportive apparatus, and the PDOP (Position Dilution of Precision) value is made possible. Figure 3 below depicts how the Altitude component describes them.

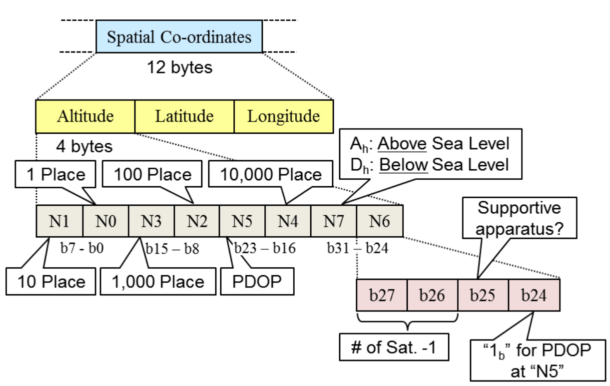


FIGURE. 3: ALTITUDE COMPONENT UPDATED IN 2004.

Before this update, an altitude was represented only as an 8-digit binary coded decimal number[2] up to 99,999,999 meters from the earth center. After this update, the altitude can be represented not only as the value from the earth center but also as a value relative to the sea level when the most significant nibble or "N7" (the upper 4 bits of the rightmost byte) is used for a sign of the value.

In addition, the nibbles "N6" and "N5" are assigned to describe the supplemental information and the PDOP value, respectively. More specifically, the most significant 2 bits of the nibble "N6" ("b27" and "b26") are assigned to represent "the number of satellites - 1", while the remaining 2 bits in the nibble or "b25" and "b24" are used as flags indicating uses of the supporting apparatus and the subsequent nibble or "N5" for the PDOP value, respectively.

Consequently, when the GPS measurement is conducted by using 4 satellites (the minimum number of satellites to determine the time and position at once) without a supporting apparatus and the PDOP value is obtained during the measurement, a value 'D_h' ($=11_b0_b1_b$) is set to the nibble "N6", resulting in expressing from -99,999 to 99.999 meter for the altitude.

Thanks to Sony HDCAM series [6] which adopted the UMID, this UMID functions were widely penetrated over the industry in the traditional tape/SDI based HD (High Definition) production environment.

Unfortunately, however, there was eventually no practical use of the UMID at that time because of awkward handling of an offline *physical* material (recorded on a tape) on the network as well as unavailability of geographical map information (there was no "Google Map" at that time!).

II. UMID for Modern Mobile News Gathering

In the modern file-based media production environment, a material as a media file is made accessible anytime and anywhere online over the network. Furthermore, a use of geographical map has already become a common practice even in the CE (Consumer Electronics) applications.

In such circumstances, time has finally come for the UMID to be effectively utilized as originally intended.

There is however a problem for the UMID descriptive power, *i.e.*, it has revealed that a camera shooting direction is much more useful than the supplemental information introduced in the previous update in order to search for a desired material, because it obviously determines whether a target object was actually shot in the material. However, having made a rich update in 2004, there is little value space remaining in the Altitude component.

To address the problem, a decision is made in SMPTE to further update the UMID format specification so that it can describe the camera shooting direction instead of the supplemental information, while still fully preserving the backward compatibility with the one for the previous update.

Figure 4 shows a structure of the Altitude component under consideration in the latest update. As discussed before,

[2] As the little endian is adopted for the UMID byte string, the leftmost byte represents the least significant 2 digits of an altitude.

the nibble "N6" is used to describe the supplemental information in the previous update. A careful analysis has revealed that the nibble "N6" can never take a value such as '3_h' (=$00_b1_b1_b$) which means that only 1 satellite is used for the GPS measurement (with a help of a supporting apparatus) while obtaining the PDOP value, because at least 3 satellites are required to obtain the PDOP value[3].

Therefore, the value of '3_h' is assigned to describe the camera shooting direction, *i.e.*, when the nibble "N6" is set to '3_h', the nibbles "N5" and "N4" are used to describe a camera shooting azimuth and a camera shooting elevation angle, respectively, as shown in Figure 4 below,

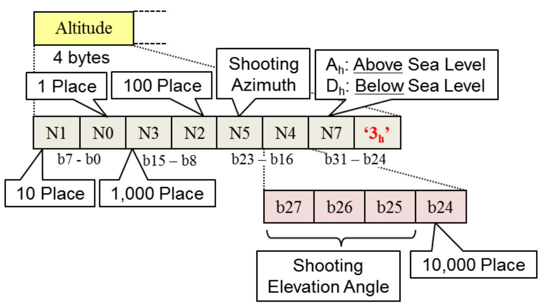


FIGURE. 4: ALTITUDE COMPONENT TO BE UPDATED IN 2017.

with the value assignment for the camera shooting direction (a shooting azimuth and a shooting elevation angles) to the nibbles "N5" and "N4" as Figure 5 below specifies.

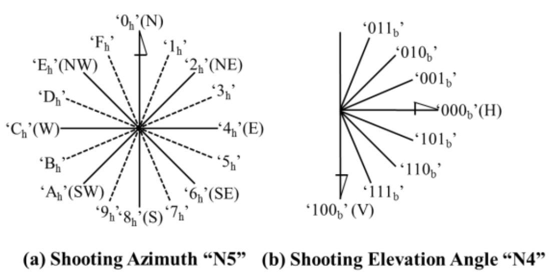


FIGURE. 5: VALUE ASSIGNMENT FOR CAMERA SHOOTING DIRECTION.

Note that the least significant bit of the nibble "N4" or "b24" is assigned to represent the most significant digit of the altitude relative to the sea level, resulting in expressing from -19,999 to 19,999 meter for the altitude.

Considering that the height of the world's highest mountain Everest is 8,848 meters and the depth of the world's deepest sea Mariana Trench is 10,911 meters, this value range is sufficient in most practical applications.

UMID APPLICATIONS

I. Shooting Direction based Material Search

Figure 6 demonstrates how the UMID is used to find a material in which a target object was actually shot. In this figure, a 3-D orthogonal coordinate system is defined by setting a camera position as the origin O, x-axis, y-axis and z-axis in the North (N), East (E) and the upward directions, respectively. Let a position of a drone as a target object (*A*) and a straight line for the camera shooting direction (*L*) at time t_0 be (x_a, y_a, z_a) and $(v_x l, v_y l, v_z l)$, respectively, in which (v_x, v_y, v_z) is a unit vector indicating the shooting direction. A distance between the point *A* and the line *L* (*d*) is defined as a length of |*AB*|, where *B* is an intersection at which a perpendicular is dropped from the point *A* to the line *L*.

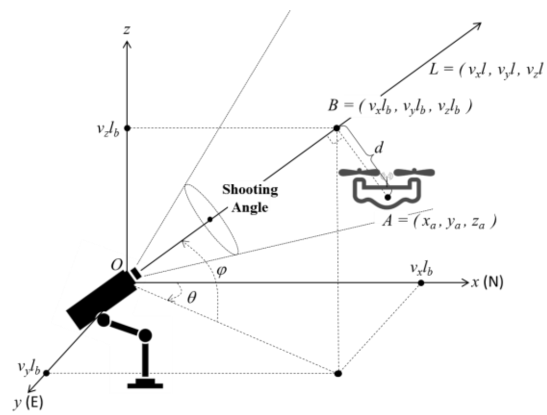


FIGURE. 6: MATERIAL SEARCH USING CAMERA SHOOTING DIRECTION.

From those notations, the distance *d* is calculated by

$$d^2 = (x_a^2 + y_a^2 + z_a^2) - l_b^2, \quad (1)$$

where l_b is a constant value given by

$$l_b = v_x x_a + v_y y_a + v_z z_a \,. \quad (2)$$

Whether a target object was shot by a material in question at time t_0 is determined by a comparison between the distance *d* and a threshold D_{th} predetermined according to the camera shooting angle at that time[4], *i.e.*, if $d < D_{th}$ is hold, then the target object was likely to be taken in a frame created at time t_0 in the material.

By using the shooting azimuth θ and shooting elevation angle φ, each component of the shooting direction unit vector is expressed as

$$\begin{aligned} v_x &= \cos\varphi \cdot \cos\theta \\ v_y &= \cos\varphi \cdot \sin\theta \\ v_z &= \sin\varphi\,, \end{aligned} \quad (3)$$

where θ and φ are obtained according to Figure 5 (*e.g.*, a value '2_h' means 45 degrees for the shooting azimuth). Because the time t_0 is obtained by the "When" part of the UMID attached to a frame in question, the distance *d* can be calculated without ambiguity when the position *A* is known.

Note that while a fixed camera is assumed in Figure 6, this argument is applicable even to a moving camera when the positon *A* is considered as that relative to the camera

[3] Because the same argument is held also for values of "5_h" and "7_h" at the nibble "N6", the assignment of those values to such as the 360 degree shooting is under consideration in the latest update.

[4] This is typically 20-40 % of a distance between the camera and the object.

position whose time-dependence is described by the "Where" part of the UMID.

II. Automatic Tracking Shooting

When the positon A of a target object is made known in real-time, an automatic tracking to shoot the target object is easily realized by using UMIDs interleaved in a media stream. Figure 7 demonstrates such an application,

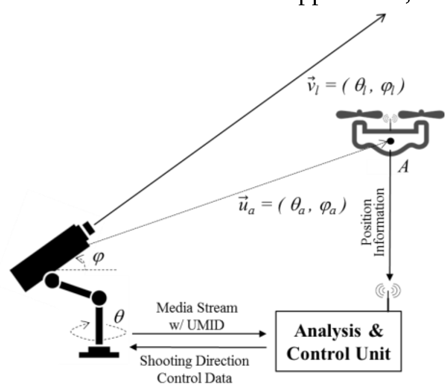


FIGURE. 7: AUTOMATIC TRACKING SHOOTING.

where an analysis and control unit (A&C Unit) is provided to receive the information on the position of a target object and on the camera shooting direction described by the UMIDs in a media stream in real-time, to analyze the information to produce unit vectors $\vec{u}_a = (\varphi_a, \theta_a)$ and $\vec{v}_l = (\varphi_l, \theta_l)$ indicating the directions toward the position A and of the camera shooting, respectively, and to generate the shooting direction control data which is then supplied to the camera to be controlled.

By controlling the camera shooting direction so that $\vec{v}_l$ (the shooting direction unit vector) matches $\vec{u}_a$ (the unit vector toward the position A), a camera can automatically tracks a target object to shoot it.

OTHER RELEVANT TOPICS IN SMPTE

I. Video over Internet Protocol (VoIP)

One of the hottest topics in the industry for now is the Video over Internet Protocol (VoIP) technology that can substitute the conventional SDI. Various technologies have been proposed, competition and collaboration are spreading among them, and yet we cannot see which will be dominant.

In SMPTE, a project is ongoing to develop a set of standards specifying the carriage, synchronization, and description of separate elementary essence streams over IP, some of which will be published in the year 2017 [7].

From the mobile news gathering application viewpoint, there might be a concern on what happens to a treatment of the ancillary information items such as the UMID. But this is not a concern.

As the SMPTE ST 291 [5] is an established and widely accepted scheme for the ancillary information items in the industry, there is an activity in IETF (Internet Engineering Task Force) to develop a standard specifying the carriage of ST 291 based ancillary data over IP [8]. Because it is positioned on the common standardized layer independent of specific VoIP proposals, the SMPTE ST 291 ancillary information items (and so the UMID) will be supported regardless of whichever proposal becomes dominant.

II. H.265/HEVC mapping to MXF

When a frame in a media stream is stored to create an MXF file, it is usually compressed to minimize its data size. How to store the compressed frame depends on the compression scheme and standards specifying this for MPEG-2 and H.264/AVC have been already provided [9].

With an appearance of a state-of-the-art compression technology H.265/HEVC, it is natural to have a preliminary discussion for it, which is just happening in SMPTE.

Again, though it might be a concern on the treatment of the ancillary information items in an MXF file for H.265/HEVC, this is not a concern, either. This is because of a provision of the MXF Generic Container [10] which specifies how to store the ST 291 based ancillary information items into an MXF file regardless of the compression scheme.

CONCLUSIONS

To meet a growing demand on the industry standard basic metadata items that can be automatically obtained during acquisition in the mobile news gathering, the SMPTE UMID is evolving to describe not only the conventional "When/Where/Who has created each frame" but also the camera shooting direction.

Although the UMID by itself is not a new technology, thanks to its history in the industry, the UMID has been well incorporated into the SMPTE standardization technology architecture. As a result, its effectiveness will always be preserved regardless of how technologies evolve including media transmission or the compression type being utilized.

REFERENCES

[1] SMPTE ST 330, "Unique Material Identifier (UMID)"

[2] SMPTE RP 205, "Application of Unique Material Identifiers in Production and Broadcast Environments"

[3] EBU Technical Review, "EBU/SMPTE Task Force for Harmonized Standards for the Exchange of Programme Material as Bitstreams", Final Report: Analyses and Results (August 1998)

[4] SMPTE ST 377, "Material Exchange Format (MXF) — File Format Specification", etc.

[5] SMPTE ST 291, "Ancillary Data Packet and Space Formatting"

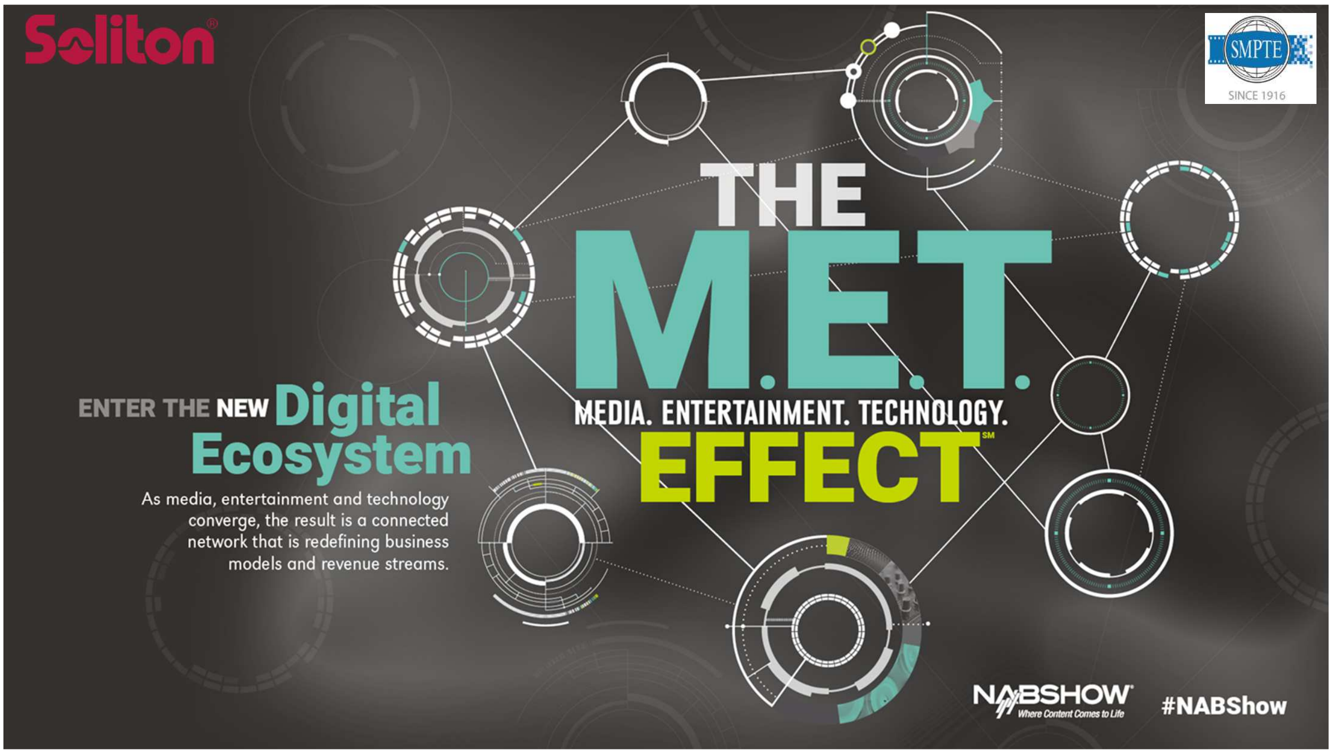
[6] E.g., https://en.wikipedia.org/wiki/HDCAM

[7] SMPTE ST 2110 series

[8] IETF Internet-draft, "RTP Payload for SMPTE ST 291 Ancillary Data"

[9] SMPTE ST 381 series

[10] SMPTE ST 379, "Material Exchange Format (MXF) — MXF Generic Container"



Soliton

SMPTE
SINCE 1916

To Maximize Interoperability in Mobile News Gathering

Yoshiaki Shibata† and Nobuo Kamata

Soliton Systems K. K.

† Chair, SMPTE TC-30MR Study Group on UMID Applications

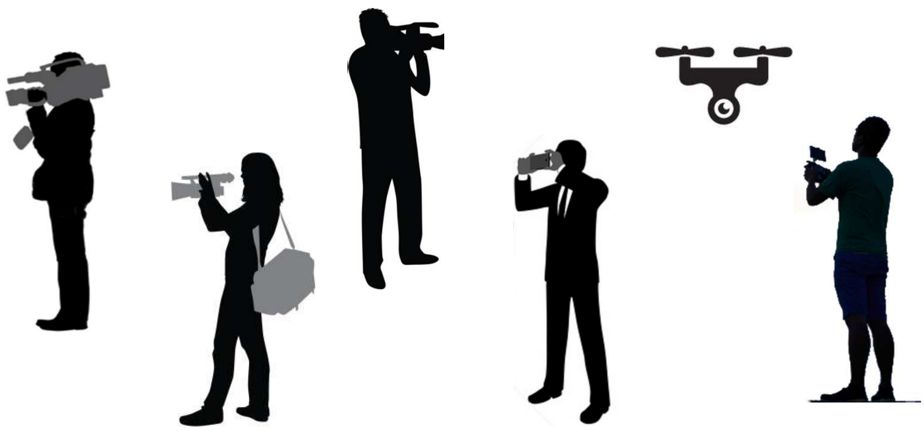
NABSHOW Where Content Comes to Life #NABShow

Outline

- Background and History
- What is the UMID?
- UMID Applications
- Relevant activities in SMPTE
- Conclusions

Modern Mobile News Gathering

- Camera shooting anytime/anywhere with a much higher mobility



#NABShow

Soliton

Audiovisual Materials and GPS Info

- Already common practice for the CE applications

(by SHUEWE)

“Geo Picture Map”

routeshoot

“RouteShoot”

What happens to the Professionals?

Soliton

In fact, … (Brief History for “File-based”)

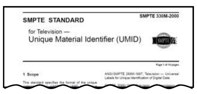
- SMPTE has already provided standard technology for it

“Origin” ***“Standardized”*** ***“Implemented”***

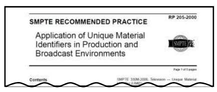

EBU Technical Review
Special Supplement 1998
EBU / SMPTE Task Force for Harmonized Standards for the Exchange of Programme Material as Bitstreams
Final Report: Analyses and Results
August 1998

EBU/SMPTE Task Force Final Report (1998)

UMID


SMPTE STANDARD
for Television — Unique Material Identifier (UMID)

SMPTE 330M (2000)


SMPTE RECOMMENDED PRACTICE
Application of Unique Material Identifiers in Production and Broadcast Environments

SMPTE RP205 (2000)

MXF


SMPTE STANDARD
Material Exchange Format (MXF) — File Format Specification

SMPTE 377M (2004), etc.

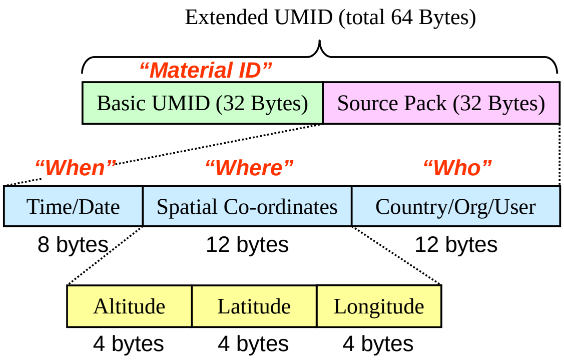
Sony HDCAM Series (1997)

Sony XDCAM Series (2004)

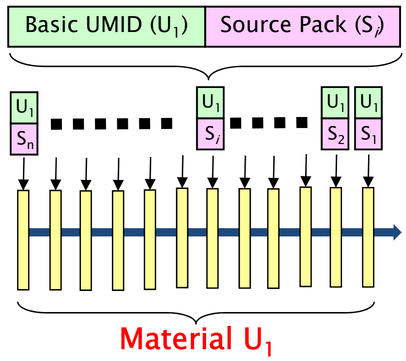
Soliton

Unique Material IDentifier (UMID)

UMID Format



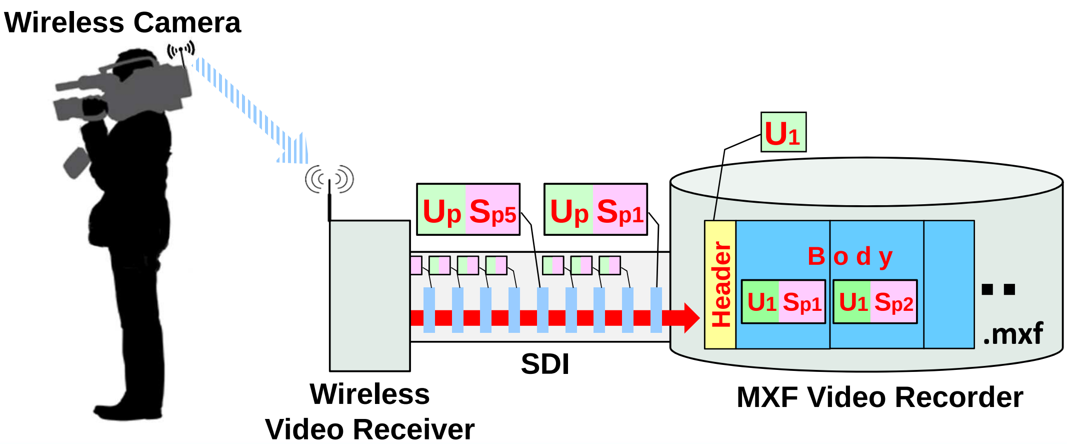
UMID for Audiovisual Material



Soliton

Mobile Acquisition with UMID

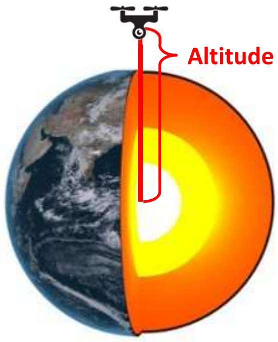
- UMID to be created automatically and stored within an MXF file



UMID Evolutions (1/2)

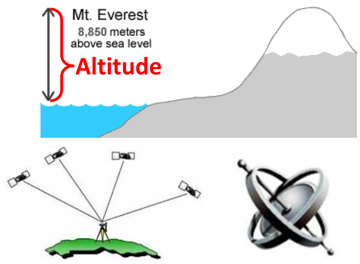
- Under evolution for “camera shooting direction” in 2017

In 2000



“Altitude” from
the center of Earth

In 2004



“Altitude” above Sea level
# of Satellites & PDOP‡ value
Supporting apparatus

In 2017

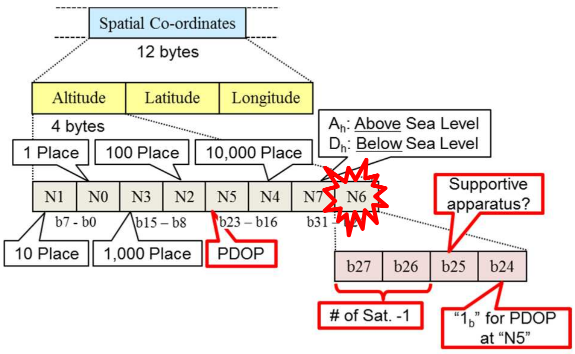
Camera shooting direction

‡ Position Dilution Of Precision

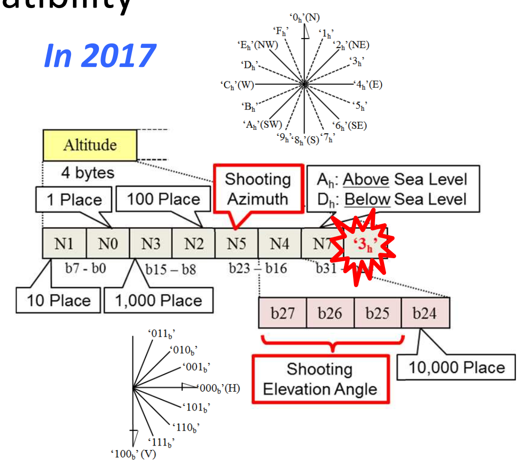
UMID Evolutions (2/2)

- To guarantee the backward compatibility

In 2004



In 2017

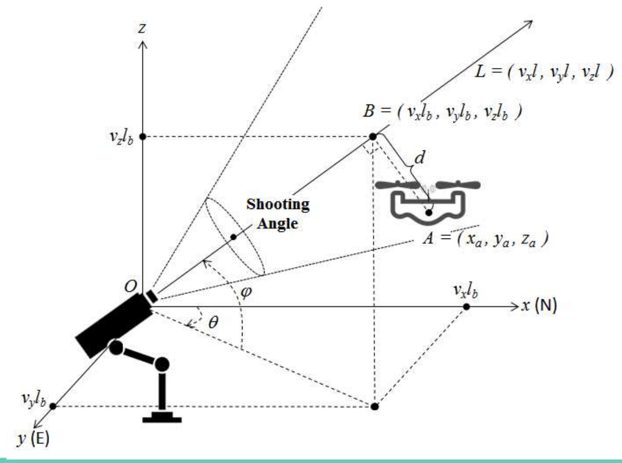


Soliton®

SMPTE SINCE 1916

UMID based Material Search

- To find materials that actually capture a target object



Distance from camera shooting direction

$$d^2 = (x_a^2 + y_a^2 + z_a^2) - l_b^2$$

where

$$l_b = v_x x_a + v_y y_a + v_z z_a$$

$$v_x = \cos\varphi \cdot \cos\theta$$

$$v_y = \cos\varphi \cdot \sin\theta$$

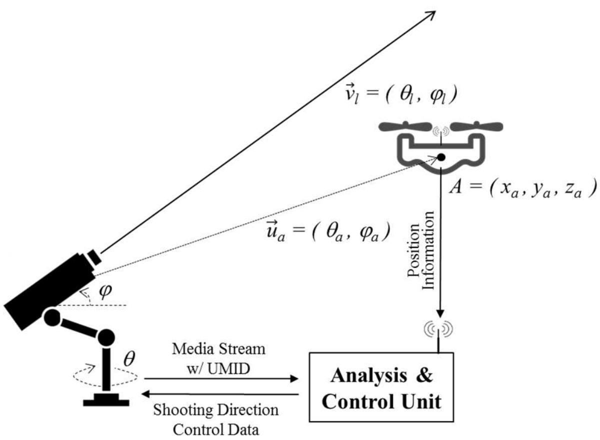
$$v_z = \sin\varphi$$

Soliton®

SMPTE SINCE 1916

UMID based Object Tracking

- Camera to automatically track a target object



Unit vector toward a target object

$$\theta_a = -\cos^{-1}\frac{x_a}{\sqrt{x_a^2 + y_a^2}}$$

$$\varphi_a = \sin^{-1}\frac{z_a}{\sqrt{x_a^2 + y_a^2 + z_a^2}}$$

to minimize

$$\Delta\theta = \theta_a - \theta_l$$

$$\Delta\varphi = \varphi_a - \varphi_l$$

Soliton

Other Relevant Topics in SMPTE

- Video over Internet Protocol (VoIP)
  - One of the hottest topics in the M&E industry
  - "RTP Payload for SMPTE ST 291 Anc. Data" by IETF addresses it
- H.265/HEVC mapping to MXF
  - Preliminary discussion is ongoing in SMPTE
  - "MXF Generic Container" by SMPTE addresses it

Soliton

To take this opportunity …

- Smart-Telecaster™ (STC) has **a potential power** to evolve your camera as "**metadata ready**"

Will **generate** and **insert** **metadata** into **video stream** at its transfer over the public network

Smart-telecaster Zao

The world's first H.265 hardware encoder

What would you expect from the future STC?

(Please drop by at "SU 4910")

Conclusions

- The higher the mobility in camera shooting, the higher the GPS infomation value!
- SMPTE UMID is ready for maximizing interoperability in mobile news gathering
- For more discussion?

See you at "SU 4910"!!

#NABShow