

MOTION IMAGERY STANDARDS PROFILE

Motion Imagery Standards Board



MISP-2015.1: Motion Imagery Handbook

October 2014

Table of Contents

Change Log	4
Scope	5
Organization	5
Chapter 1	6
Terminology and Definitions.....	6
1.1 Motion Imagery.....	6
1.1.1 Time	6
1.1.2 Frame.....	6
1.1.3 Image.....	9
1.1.4 Multiple Images.....	10
1.1.5 Full Motion Video (FMV).....	11
1.2 Metadata.....	12
Chapter 2	13
Motion Imagery Functional Model.....	13
2.1 Introduction	13
2.2 Scene	14
2.3 Imager	14
2.3.1 Imager Processing Model.....	14
2.3.2 Metadata.....	16
2.4 Platform.....	16
2.5 Control.....	16
2.6 Exploitation	16
2.7 Archive	17
2.8 Building Block Functions.....	17
2.8.1 Compression.....	17
2.8.2 Encodings.....	17
2.8.3 Protocols.....	17
2.8.4 Processing	18
Chapter 3	19
Imager Types	19
3.1 Direct Imager.....	19
3.1.1 Global Shutter	19
3.1.2 Rolling Shutter	20
3.1.3 Interlaced.....	21
Chapter 4	22
Image Color Model.....	22
Chapter 5	24
Dissemination	24
5.1 Background	24
5.1.1 Transmission Methods	24
5.1.2 Internet Protocols	25
Appendix A References	29
Appendix B Acronyms	30

List of Figures

Figure 1-1: Samples, Pixels, Bands and Frame.....	7
Figure 1-2: Generation of a Frame.....	8
Figure 1-3: Image is a subset of Frame.....	9
Figure 1-4: Example of Spatial Overlap	10
Figure 1-5: Relationships: Frame-to-Video and Image-to-Motion Imagery.....	11
Figure 2-1: Elements of the Motion Imagery Functional Model	13
Figure 2-2: Motion Imagery from a Varieties of Modalities	15
Figure 2-3: Imager Processing Model.....	15
Figure 3-1: Types of Imagers.....	19
Figure 3-2: Example Motion Effects: Global vs. Rolling Shutter	20
Figure 4-1: Examples of Formats with Chroma Subsampling.....	22

List of Tables

Table 4-1: Pixel Value Range for Various Color Sampling Formats	23
Table 5-1: Internet Protocols	25
Table 5-2: UDP Error Types.....	26
Table 5-3: MPEG-2 TS Error Types.....	27

Change Log

Scope

The purpose of the Motion Imagery Handbook is to provide:

1. A definition of Motion Imagery.
2. Common terminology for all MISB documentation.
 - a. There is no single authoritative source for technical definitions of terms within the community; therefore, the Motion Imagery Handbook serves as the authoritative source of definitions for the MISB community of practice.
3. Additional detail for topics identified in the Motion Imagery Standards Profile [1].
 - a. The MISP succinctly states requirements, while the Motion Imagery Handbook discusses principles underlying requirements more thoroughly.

Although intended to be educational and informative the Motion Imagery Handbook is not a substitute for available material that addresses the theory of imaging, video/compression fundamentals, and transmission principles.

Organization

The Motion Imagery Handbook is composed of chapters, each emphasizing different topics that support the Motion Imagery Standards Profile (MISP). Each chapter is intended to be self-contained with references to other chapters where needed. Thus, a reader will be able to quickly locate information without reading preceding chapters. The Motion Imagery Handbook is expected to mature over time to include material considered essential in applying the requirements within the MISP as well as other MISB standards.

Chapter 1

Terminology and Definitions

1.1 Motion Imagery

Many different sensor technologies produce Motion Imagery. To support an imaging workflow in which sensor data can be utilized by an Exploitation system, standards defining common formats and protocols are needed. The standards facilitate interoperable functionality, where different vendor products can readily be inserted within the workflow based on improvement and cost. Such standards need to be developed on a well-defined and integrated system of terminology. This section lays the foundation for this terminology.

1.1.1 Time

Time is fundamental in Motion Imagery. All events whether captured by a camera or artificially created are either formed over a period of time, or are displayed over a period of time. In a camera, for instance, the light reflected from an object is exposed onto the cameras “sensor”, which could be some type of imaging array or film. The period of exposure is bounded by a **Start Time** and an **End Time**. These are important qualifiers that play a significant role in image quality, particularly in capturing motion and the signal to noise ratio.

Time can be absolute or relative. Although these terms have various definitions in the literature, here they are defined specific to their application in the MISB community. **Absolute Time** is measured as an offset to a known universal source, such as Coordinated Universal Time (UTC). **Relative Time** is measured as an offset from some starting event. For example, Relative Time is a basis for overall system timing in the formatting of Motion Imagery, compression and data transmission.

Start Time: Time at which a process is initiated, measured in either Absolute Time or Relative Time.

End Time: Time at which a process is completed, measured in either Absolute Time or Relative Time.

Absolute Time: Time that is measured as an offset to a known universal source’s (e.g. UTC) starting point.

Relative Time: Time that is measured as an offset from a starting event.

1.1.2 Frame

The term Frame is commonly used in describing video and Motion Imagery, for instance, *image frame*, *frame size*, *frames per second*, etc. A **Frame** is defined as a two-dimensional array of regularly spaced values, called Pixels that represent some type of data – usually visual.

A **Pixel** is a *combination* of one or more individual numerical values, where each value is called a **Sample**. A **Sample** is data that represents a measured phenomenon such as light intensity.

In considering visual information, a **Frame** could be the data representing a black and white or color picture, for example. In a black and white picture, the **Pixel** values are the intensity values at each position in the picture mapped into the **Frame**. In a color picture, the **Pixel** data is composed of three different intensity values, i.e. red, green and blue at each position in the picture.

An array of **Samples** where all phenomena are of the same type is called a **Band**. For example, a **Band** of **Samples** for the red component of color imagery contains only measurements of light sensitive to the “red” wavelength. For a black and white picture, one **Band** is sufficient, whereas for color three **Bands** are needed. A **Frame** can consist of **Pixels** combined from one or more **Bands**. Figure 1-1 illustrates the relationships of **Samples** and **Bands** to a **Frame**.

A **Pixel** is a *combination* of a number of **Samples** collectively taken from a number of **Bands** (see Figure 1-1). Where there is only one **Band**, a **Pixel** is equivalent to a **Sample**. A “color” **Pixel** is a combination of red, green and blue **Samples** from corresponding red, green and blue **Bands**. In Chapter 4 various color models where the relationship between **Pixels** and **Samples** are not one-for-one are discussed.

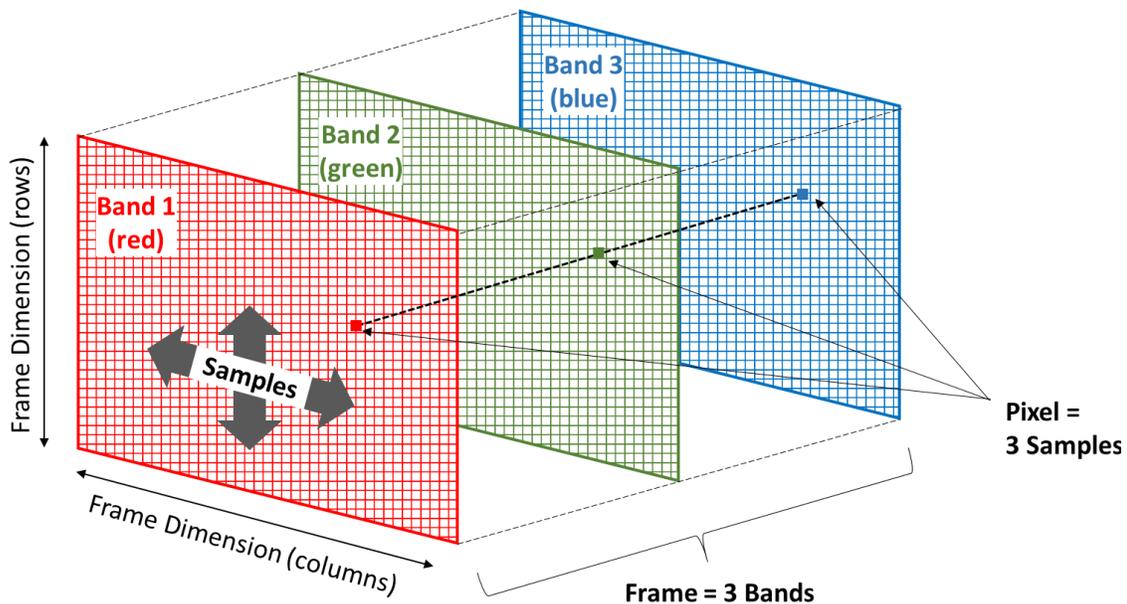


Figure 1-1: Samples, Pixels, Bands and Frame

The **Frame Dimension** is the height of the **Frame** measured in **Pixels** per column and the width measured in **Pixels** per row. **Pixels** within a **Frame** are bounded in their measurement over a period of time; that is, they have a **Start Time** and **End Time**, which is typically called the period of integration. The **Start Time** and **End Time** may be based on **Relative Time** or **Absolute Time**. Although all **Pixels** within a **Frame** generally have the same **Start Time** and **End Time**, this is not always the case. A **Frame** is bounded by a **Frame Start Time** and **Frame End Time**, which accounts for the extremes of **Start Time** and **End Time** for the individual **Pixels**.

Sample: a numerical value that represents measured phenomena, such as light intensity along with its Start Time and End Time.

Band: collection of Samples where all measured phenomena are of the same type.

Pixel: A combination of a number of Samples collectively taken from a number of Bands.

Frame: A two-dimensional array of regularly spaced Pixels in the shape of a rectangle indexed by rows and columns along with a Start Time and an End Time of each Pixel.

Frame Dimension: The height and width of a Frame measured in Pixels per column and Pixels per row, respectively.

Frame Start Time: The minimum time value of all Pixel Start Times within a Frame.

Frame End Time: The maximum of all Pixel End Times within a Frame.

In generating a Frame of measured phenomena, Source Data (the sensed phenomena) is further conditioned by some Data Source Processing as illustrated in Figure 1-2.

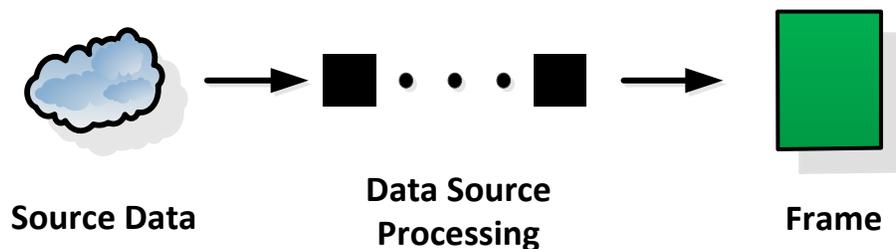


Figure 1-2: Generation of a Frame

The Source Data may be Visible Light, Infrared, a LIDAR cloud, RADAR returns, Acoustical or any type of data from any type of **Modality**. Source Data can be computer generated or data from the result of a simulation or other processing. The Data Source Processing maps the Source Data to a two-dimensional rectangular output – the Frame. The Data Source Processing depends on the type of Source Data; it can be integral to the sensor, or exist as a separate system.

Modality: The type of Source Data (e.g. Visible Light, IR, LIDAR, etc. or combination of these).

Examples of Data Source Processing include: Visible Light or Infrared (IR) cameras; the post processing of a 3D LIDAR cloud that supports a viewport into the 3D LIDAR scene; a simulation of flying over a city; simple text annotation. The Data Source Processing can provide contextual Metadata for the Frame and how it was formed.

The Data Source Processing may produce near-instantaneous Frames, where all Samples or Pixels are captured at the same time, or Frames where the data is integrated over a period a time. Both types of Frames are bounded with a Frame Start Time and a Frame End Time (for the near-instantaneous case the Frame Start Time and Frame End Time are considered identical).

1.1.3 Image

A Frame can represent content from any Source Data; this data can be a constant value, computer generated graphics, virtual reality, or real world forms of energy. An **Image** is a subset of all possible Frames created from sensor Source Data. Sensor Source Data is data from any device that detects information from the physical world. The space in the physical world imaged by the Sensor is called the **Scene**. Examples of Sensor Source Data are Visible Light, Infrared, LIDAR Cloud and SAR data. Examples of data that are NOT Sensor Source Data are text, computer graphics, maps, and title slides.

The definition of Image, which is the result of processing some Source Data collected by a sensor, is built upon the definition of Frame.

Image: A Frame with Pixels derived from sensed phenomena.

Scene: Space in the physical world that is sensed by a sensor and used to form an Image.

Image Dimension: The height of an Image measured in Pixels per column and the width of an Image measured in Pixels per row.

Image Start Time: The Start Time of an Image.

Image End Time: The End Time of an Image

Newscast graphics and computer animation are Frames, but because they are not produced from sensor data they are not Images. In contrast, the pictures from an air vehicle sensor, underwater sensor and sonar sensor are all Images, because they are formed from sensed data. Image is a subset of Frame (depicted as in Figure 1-3), therefore, Images retain all of the attributes of Frames (i.e. rectangular array structure and time information).

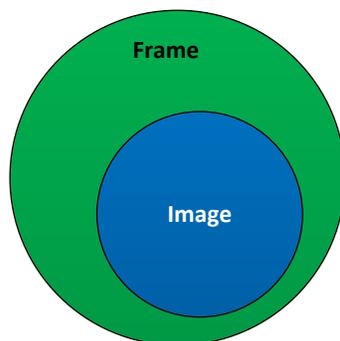


Figure 1-3: Image is a subset of Frame

1.1.4 Multiple Images

With two or more Images, relationships amongst Images can be formed both spatially and temporally. When two Images contain some portion of the same Scene, there is spatial overlap – these are called **Spatially Related Images**. Figure 1-4 illustrates spatial overlap, where the red square outlines similar content in each Image. Spatially Related Images do not necessarily need to occur within some given time period. For example, the two Images in Figure 1-4 may have been taken within milliseconds, minutes, or hours of one another. Spatially Related Images may be separated by a large difference in time, such as Images of a city taken years apart.



Figure 1-4: Example of Spatial Overlap

Images collected at some regular time interval, where the Images form a sequence, the Image Start Time for each is known, and each successive Image temporally follows the previous one, are called **Temporally Related Images**. There is no criteria that the content within Temporally Related Images be similar, only that they maintain some known time relationship.

Spatiotemporal data is information relating both space and time. For example, capturing a scene changing over time requires a sequence of Images to be captured at a periodic rate. While each Image portrays the spatial information of the scene, the sequence of these Images portrays the temporal or time-varying information. Images that are both spatially and temporally related are called **Spatio-Temporally Related Images**. These are the type of Images found in Motion Imagery.

Spatially Related Images: Images where recognizable content of a first Image is contained in a second Image.

Temporally Related Images: When the two Image Start Times of two Images are known relative to each other; the second Image is always temporally after the first.

Spatio-Temporal Related Images: When two Images are both Spatially Related Images and Temporally Related Images.

By collecting a series of Frames and/or Images, **Video** and **Motion Imagery** can be defined. The term “video” is not well defined in the literature. The word video is Latin for “I see.” It has become synonymous with standards and technologies offered by the commercial broadcast

industry. As such, the term serves a rather narrow segment of the application space served by Motion Imagery.

Video: An ordered series of Frames with each Frame assigned an increasing Presentation Time; where the Presentation Time is a Relative Time.

Presentation Time: A Relative Time associated with each Frame.

This definition of Video includes Presentation Time, which is an arbitrary relative timeline that is independent of a Frame's Start Time. For example, a Video of a glacier ice flow created by taking one picture per day has a Frame Start Time that is 24 hours apart from the next Frame; the Presentation Time, however, is set to play each Frame at a 1/30 second rate (i.e. Video at 30 frames per second).

Motion Imagery: A Video consisting of Spatio-Temporally Related Images where each Image in the Video is spatio-temporally related to the next Image.

Video is created from a sequence of Frames, whereas Motion Imagery is created from a sequence of Spatio-Temporal Related Images. Just as Image is a subset of Frame, Motion Imagery is a subset of Video. These relationships are depicted in Figure 1-5.

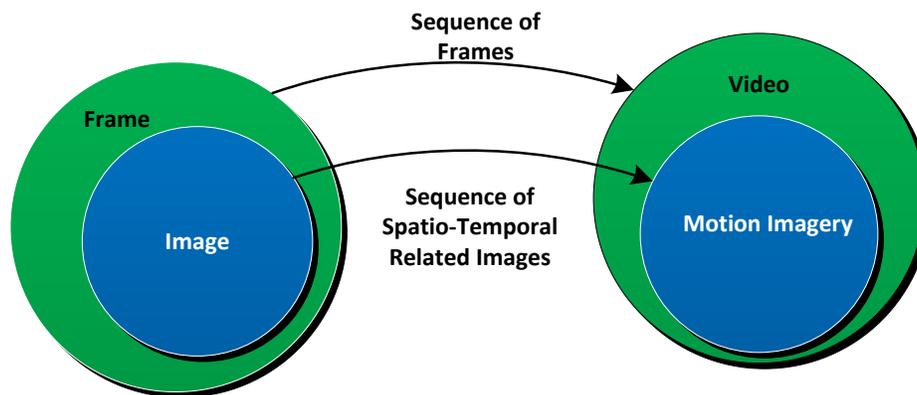


Figure 1-5: Relationships: Frame-to-Video and Image-to-Motion Imagery

1.1.5 Full Motion Video (FMV)

The term **Full Motion Video (FMV)** loosely characterizes Motion Imagery from Visible Light or Infrared sensors, playback rates typical of Video, and Frame Dimensions typical of those found in the commercial broadcast industry, defined by standards development organizations like SMPTE and ISO.

As with video, the term FMV characterizes a rather narrow subset of Motion Imagery. It is recommended the term FMV not be used, because of its ill-defined and limited applicability across the diverse application space served by Motion Imagery. Moreover, there is no clear definition for FMV available – it is sort of tribal knowledge and varies depending on who is asked. Historically, the term FMV was coined in the 90's by a vendor of video transponders to describe analog video that could be played back at its native frame rate showing all of the “motion” in the video.

The term FMV should not be used in contractual language.

1.2 Metadata

Motion Imagery is the *visual* information that is exploited; however, in order to evaluate and understand the context of the Motion Imagery and its supporting system additional information called **Metadata** is needed. The types of Metadata include information about the sensor, the platform, its position, the Image space, any transformations to the imagery, time, Image quality and archival information. Many MISB standards specifically address the definition of Metadata elements and the formatting of the Metadata associated with Motion Imagery.

Chapter 2

Motion Imagery Functional Model

2.1 Introduction

A Motion Imagery Functional Model offers a common narrative across different audiences, such as Program Managers/Procurement officers, Technical Developers and End Users. The Functional Model describes the elements of systems that generate, manipulate and use Motion Imagery, and is based on the logical data flow from the Scene to the Analyst as shown in Figure 2-1. These elements include:

- 1) Scene - the data source for the Motion Imagery
- 2) Imager - a Sensor or Processor that converts Scene data into Images
- 3) Platform - static or movable system to which the Imager is attached
- 4) Control - a device that directs the Imager position, orientation or other attributes
- 5) Exploitation - the human/machine interaction with the Motion Imagery
- 6) Archive - stores Motion Imagery and additional exploitation data

In addition to these elements there are processing functions (denoted in the red block of Figure 2-1) used to format and manipulate the Motion Imagery; these are also included in the Functional Model.

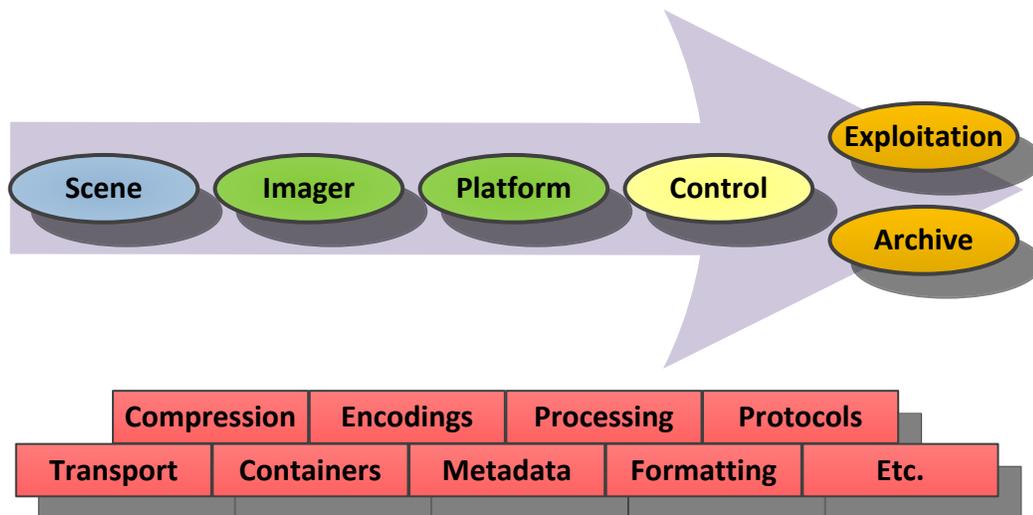


Figure 2-1: Elements of the Motion Imagery Functional Model

Using the Motion Imagery Functional Model, MISB Standards (ST) and Recommended Practices (RP) that address particular stages in the model are related. This facilitates ready association to those standards that are mandatory when specifying system requirements.

The Building Block Functions are core to MISB ST and RP documents; these may be cast in standalone documents, or as guidance provided within the MISP, where the function is defined generically and then referenced within MISB STs and RPs.

2.2 Scene

The Scene is what is being viewed by an Imager. Different modalities may be used to construct an Image of a Scene. For a given modality, the Scene is the Data Source. Each Scene may produce multiple Data Sources at the same time if multiple modalities are used. Typical modalities include:

- Electro-Optical - Emitted or reflected energy across the Visible/Infrared portion of the electromagnetic spectrum (ultraviolet, Visible, near IR, and IR).
 - Visible Light - Color or Panchromatic
 - IR Imaging - Pictorial representation of thermal IR emissions
 - Spectral Imagery - Discrete bands of the electromagnetic spectrum are imaged individually
 - MSI - Multispectral Imagery - 10's of individual bands
 - HSI - Hyperspectral Imagery - 100's of individual bands
- RADAR - Reflected energy in the radio frequency portion of the electromagnetic spectrum.
- LIDAR - Laser pulses are timed from transmitter production and reflection back to a receiver.

2.3 Imager

The Imager converts information from a Data Source into an Image, and when possible provides supporting information, such as the Imager characteristics and time about when the Samples or Pixels were created. Information that supports the Imager is called Metadata. The MISP specifies requirements on the format of imagery produced by an Imager, such as horizontal and vertical Sample/Pixel density, temporal rates, and Sample/Pixel bit depth. These requirements assure that common formats and structures are used, thereby facilitating interoperability.

2.3.1 Imager Processing Model

Figure 2-2 illustrates the variety of modalities used to create Motion Imagery. Information from a Data Source is processed into Frames along with associated Metadata.

While a modality and its subsequent processing may be unique, a common form, the Frame, is the result (plus modality-specific Metadata).

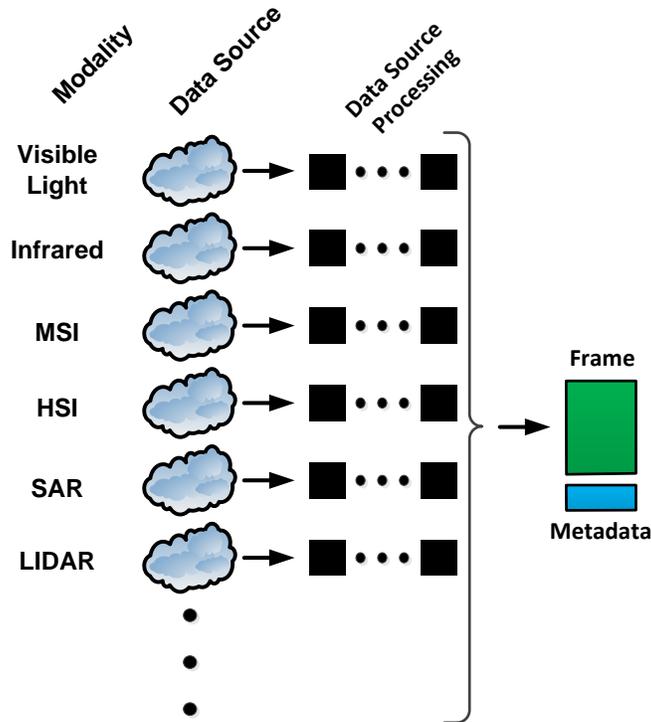


Figure 2-2: Motion Imagery from a Varieties of Modalities

There are many types of Imagers depending on the modality. Viewed at a finer level, producing an Image is a multi-step process with each step using the output of one or more previous steps spatially, temporally, or both. When performing precise exploitation of the Motion Imagery it is important to understand what data manipulations have been performed on the original Source Data. An Image Processing Model (see Figure 2-3) provides a consistent method for recording the process.

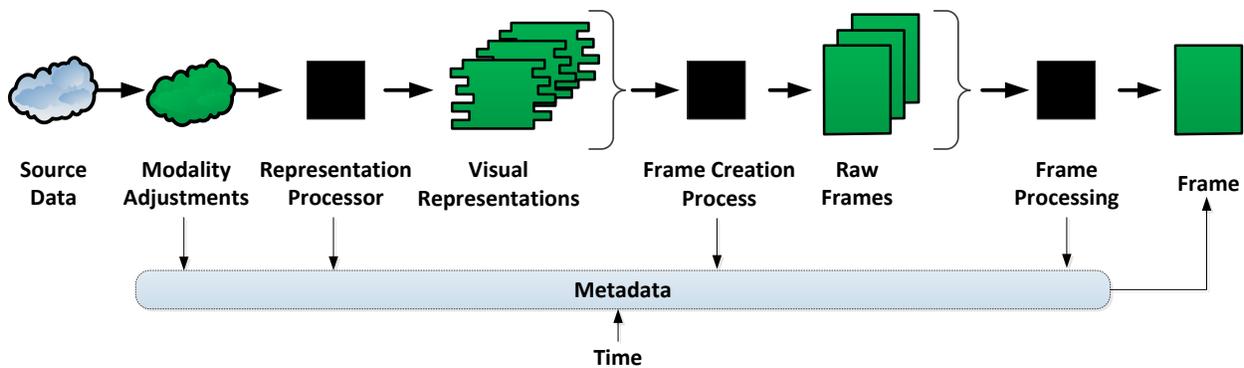


Figure 2-3: Imager Processing Model

The Imager Processing Model shows the information flow from left to right:

- **Source Data:** the raw data that emanates from a Scene for a particular modality or set of modalities. The Source Data is a collection of raw data measured from the Scene.
- **Modality Adjustments:** adjustments to the raw data that occur in the modality phenomenology; for example, atmospheric distortions and lens distortions.

- **Representation Processor:** maps the data into a two-dimensional digital representation called a Visual Representation. Examples include: CCD camera, CMOS camera, Infrared camera and LIDAR post-processing system.
- **Visual Representation:** a two-dimensional array of Samples (in any configuration i.e. shape or spacing) that are measurements from the Scene. Each Sample has a numeric value, a location relative to the other Samples, a Start Time and an End Time.
- **Frame Creation Process:** maps a Visual Representation to a set of regularly spaced Homogenous Samples or Pixels in the shape of a rectangle. The mapping is dependent on the type of modality and Visual Representation(s). Frames can be created from one or more Visual Representations either temporally or spatially.
- **Raw Frame:** same as definition of Frame in Section 1.1.1.
- **Frame Processing:** either augmentation or manipulation of one or more Frames (spatially or temporally) for the purpose of formatting the Frame to a specific size or temporal constraints.
- **Frame:** as defined in Section 1.1.1.

An important aspect of the Image Processing Model is the temporal information recorded during the Image formation process. The Image Processing Model defines a number of steps; however, depending on the modality and sensor type not all of the steps are needed (and can be skipped) to produce a Frame. The only required step is the Representation Processor.

2.3.2 Metadata

In addition to the imaged data, the orientation and position of the collected Visual Representation is needed.

2.4 Platform

Any system to which the Imager is attached may be considered its platform. A platform may provide information regarding its environment, such as time, place, condition of the platform, etc. that may be quantified and provided in the form of Metadata along with Imager essence. The MISP provides numerous Metadata elements that serve specific purposes within its suite of Standards documents.

2.5 Control

Motion Imagery systems generally allow for control over the Imager, whether orienting its direction dynamically, or modifying its parameters, such as contrast, brightness, Image format, etc. The MISB does not issue guidance for control of a platform; it does, however, prescribe Metadata to enable Image transformations whether at the platform or in later phases of processing.

2.6 Exploitation

Exploitation of Motion Imagery may range from simple situational awareness – the when and where, to in-depth extraction of detected features, measurement, and coordination with other intelligence data. Because the tools used in exploitation operate on the data structures of Motion

Imagery, revisions to the MISP are done in as backward compatible way possible, so all operational tools may continue to function as new capabilities are made available. While this is a goal, the advance and adoption of new technologies may impact compatibility in some cases.

2.7 Archive

Motion Imagery content is stored for later phases of exploitation, generating reports and historical reference for comparison. An important aspect of storage is file format. Choosing a standardized file format and a means to database/search the Motion Imagery is critical to reuse. The MISP provides guidance on several file containers, and continues to evaluate new technologies that may offer greater value in the community.

2.8 Building Block Functions

A Building Block Function is itself a MISB standard that defines a reusable function that supports other higher-level MISB standards.

2.8.1 Compression

Motion Imagery typically is output by an Imager as a number of continuous sequential Images, where each Image contains a defined number of Samples/Pixels in the horizontal direction (columns) and a defined number of Samples/Pixels in the vertical direction (rows). The Images are spaced at a fixed time period.

Compression is an algorithmic sequence of operations designed to reduce the redundancy in a Motion Imagery sequence, so the data may be transported within a prescribed bandwidth transmission channel. The tradeoffs in compressing Motion Imagery are data rate, Image quality and stream latency. These must be optimized on a per-application basis. The MISB governs the type of compression and provides guidelines for its proper use.

Audio is another “essence” type that may be provided by the platform. It also is typically compressed, and the MISB allows a choice among several industry standards.

2.8.2 Encodings

An encoding is the process of putting a sequence of characters (letters, numbers, and certain symbols) into a specialized format for efficient transmission or storage. Encodings such as KLV (Key Length Value) format are designed for low overhead representations of Metadata. While many MISB Standards assume KLV encodings for Metadata, the MISB is investigating other encodings for use in web-enabled environments.

2.8.3 Protocols

Protocols provide the linkage for systems to communicate; they are key to interoperability. Protocol include the interface specifications for data transfer between functions along the Motion Imagery Functional Model. MISB chooses protocols specified by the commercial and international standards development organizations. When appropriate these protocols are further profiled for specific use in this community, which aids interoperability and conformance.

2.8.4 Processing

Many points along the data flow within the Motion Imagery Functional Model are acted on for conversion, formatting and improvement to the signals passed. Examples include image transformations, data type conversion, and clipping of streams into smaller files. While the MISB does not dictate specific implementations of processing, the Standards are designed to provide flexibility and consistency across implementations.

Chapter 3

Imager Types

There are two types of Imagers: Direct and Indirect. A Direct Imager transforms the raw Source Data information into an Image. Examples of direct Imagers include Visible Light cameras, Infrared Cameras, and Hyperspectral Sensors that gather data, perform some processing and generate the Image from the same perspective as the sensor. An Indirect Imager transforms the raw Source Data into an intermediate form, which is then processed into an Image via projective methods. A LIDAR sensor is an example of an Indirect Imager that produces Images by first building a point cloud; Images are then built from “flying” around the point cloud. Figure 3-1 illustrates the difference between the two types of Imagers.

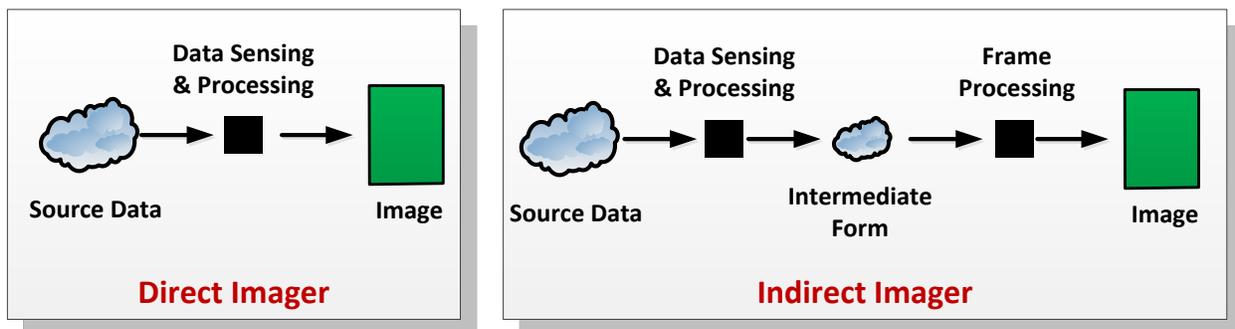


Figure 3-1: Types of Imagers

3.1 Direct Imager

There are primarily three type of direct imagers: progressive-global, progressive-rolling and interlaced.

3.1.1 Global Shutter

In Imagers that use a global shutter, all Samples or Pixels within an entire Image are “imaged” nearly simultaneously. At the Image Start Time any residual signal in the sensor elements is removed. Once the shutter is opened and the Imager array exposed to light, the Imager elements accumulate energy for some period of time. At the end of the integration period (time during which light is collected = Image End Time - Image Start Time), the energy is all Imager elements are simultaneously transferred to a light-shielded area of the Imager. The light shield prevents further accumulation of energy during the readout process. The energy in the elements is then read out and the process repeated.

With a global shutter Imager, the scene is "frozen" in time, provided the integration time is short enough i.e. there is no change in the scene during the integration time. The advantages of a global shutter is superior motion capture capability over rolling shutter Imagers. The disadvantages include higher cost and larger size.

3.1.2 Rolling Shutter

In a rolling shutter Imager the elements in the Image do not collect light at the same time. All elements in one row of the Imager do collect light during exactly the same period of time, but the time light collection starts and ends is slightly different for each row. For example, the top row of the Imager is the first one to start collecting light and is the first one to finish collecting. The start and end of the light collection for each successive row is slightly delayed. The total light collection time for each row is exactly the same, and the time difference between rows is constant.

The time between a row being set to be exposed to light and that row being read is the integration time. The advantage of a rolling shutter method is the Imager can continue to gather energy during the acquisition process, thus increasing sensitivity. The disadvantages of rolling shutter include distortion of fast-moving objects or flashes of light, or rolling shutter artifacts, which are most noticeable when imaging in extreme conditions of motion. The majority of rolling shutter technology is found in the consumer market i.e. cell phones. Since the integration process moves *through* the Image over some length of time, users should keep in mind the following issues when working with rolling shutter sensors:

3.1.2.1 Motion Blur and Wobble

Motion blur may occur depending on the speed of motion within the scene. This can severely distort the imagery to a point where objects no longer appear natural and uniform. For example, when viewing a fan the blades may take on irregular shape (see Figure 3-2). When the Imager is subject to vibration, the Image may appear to wobble unnaturally.

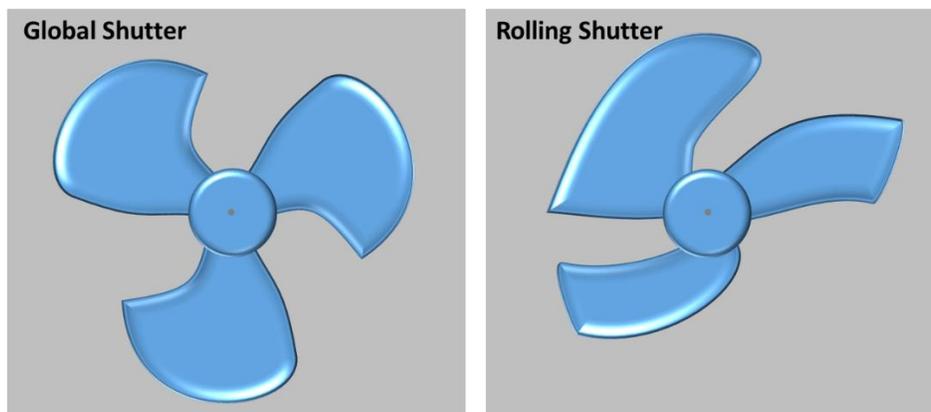


Figure 3-2: Example Motion Effects: Global vs. Rolling Shutter

3.1.2.2 Strobe Light and Partial Exposure

A rolling shutter Imager is not well-suited for capturing short-pulse light sources, such as strobe light or flash. Unless the light source remains on for the duration of exposure, there is no guarantee of adequately capturing the light source. This will result in an Image with varying levels of illumination across the scene. As this effect manifests differently in successive Images of Motion Imagery sequence, the imagery may appear to “breathe” with some content possibly washed out completely.

3.1.3 Interlaced

In an interlaced scan Imager, the Image is “imaged” in two passes: one-half of the horizontal Pixel rows are captured during the first pass, and the other half in the next. Thus, two complete passes (or scans) are required to capture one complete Image. One main drawback of interlaced scanning is that Images tend to flicker, and motion – especially vertical motion – appears jerky. A second drawback, is Image detail, like object edges, can be “torn” demonstrating a stair-step “jagged” effect along an object edge. As the motion increases stair-stepping can become quite pronounced greatly distorting Image features essential to exploitation tasks. This distortion is further compounded when compressing an Image. Because the stair-stepping artifact introduces higher frequency detail, coding efficiency is reduced as the coder attempts to spend its allocated bits representing these artifacts. With higher compression ratios, these artifacts are even further degraded.

Interlaced-scan is an older technology developed to deliver analog television. Because of the duration this technology has been in the marketplace, it is inexpensive making it attractive. However, it is a poor choice for surveillance applications where motion is common.

Chapter 4

Image Color Model

Color images are generally represented using three Bands comprised of a number of Samples per Band interpreted as coordinates in some color space. A color space is a mathematical representation of a set of colors. Several popular color spaces include RGB (Red-Green-Blue), YUV and YCrCb (where Y is the luminance, and UV and CrCb are color difference values). Color spaces such as YUV and YCrCb were developed as more efficient representations (i.e. the color difference values require less bandwidth than the luminance value) derived as linear combinations of the RGB values. Nomenclatures of 4:4:4, 4:2:2 and 4:2:0 denote spatial sampling of the color Bands. The graphic in Figure 4-1 helps to explain color sampling for these common encodings.

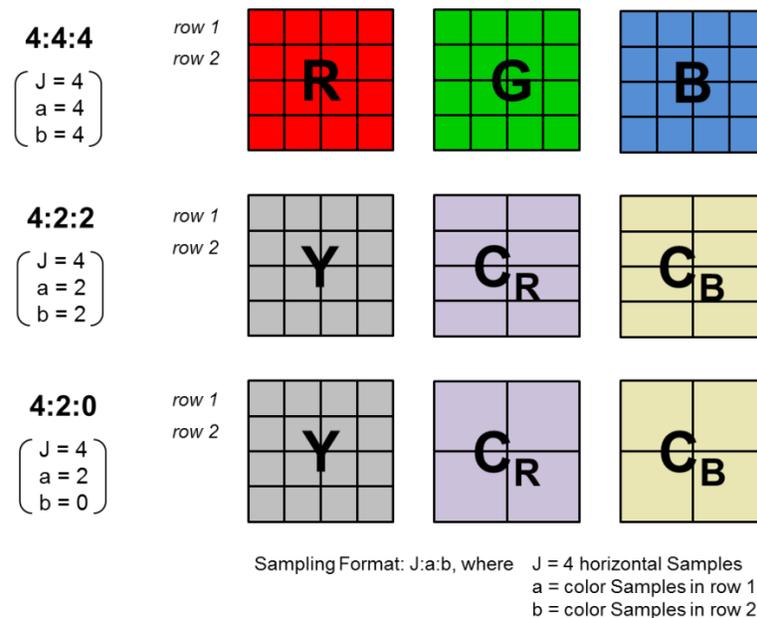


Figure 4-1: Examples of Formats with Chroma Subsampling

At the top of the figure a set of 4x4 Sample arrays represent three color Bands, one each for Red, Green and Blue. Likewise, the middle and bottom show three Sample arrays that represent the color Bands of Luminance (Y) and two Chrominance (Cr Cb). A sampling ratio with notation J:a:b is used: “J” is the dimension of the array horizontally, in this case J = 4; “a” indicates the number of Samples in row 1, and “b” the number of Samples in row 2.

For example, in 4:4:4 (top of Figure 4-1) each of the three bands have the same spatial sampling; that is, each band (RGB in the example) has a Sample that represents color information in each Pixel location. This contains the maximum number of Samples, which is 48 (16 Y+ 16 Cr + 16 Cb).

In 4:2:2 (middle of Figure 4-1), every two Samples in row 1 share a Chroma Sample ($a=2$); likewise, every two Samples row 2 share a Chroma Sample ($b=2$). For 4:2:2 when forming a Pixel, a single Chroma Sample is reused by two Pixels (the Pixel's row-wise neighbor); this reduces the number of Samples by one-third to 32 Samples ($16 Y + 8 Cr + 8 Cb$).

In 4:2:0 (bottom of Figure 4-1), every two Samples in row 1 share a Chroma Sample ($a=2$); the row 2 shares its Chroma Sample with the top row. For 4:2:0 when forming a Pixel, a single Chroma Sample is reused by four Pixels (the Pixel's row-wise and column-wise neighbors); this reduces the number of Samples by one-half to 24 Samples ($16 Y + 4 Cr + 4 Cb$).

Often a Pixel, such as "24 bit color Pixel" or "16 bit color Pixel", describes a 3-band set of Sample values. Determining the Pixel value for three bands each with the same spatial sampling is straightforward, i.e. Pixel Value Range = $3B$, where B = bits per Sample for one band. In the case of color sampling, an *equivalent* Pixel Value Range can be computed in reference to the Pixel arrangement shown in Figure 4-1. Note that in color sampling of 4:2:2 and 4:2:0 the chrominance bands have fewer Samples than the luminance band.

In Table 4-1, the Pixel Value Range in bits per Pixel for the three color samplings are listed for several Sample Value Ranges of 8, 10 and 12 bits per Sample. The Pixel Value Range is based on the number of possibly unique samples within the sample array. For instance, 4:4:4 has equal sample spacing in each band, so there is one Sample in each band, i.e. full sample density. In 4:2:2 for every one Sample in band 1 there are 0.5 Samples in band 2 and 0.5 Samples in band 3. Likewise, in 4:2:0 for every one Sample in band 1 there are 0.25 Samples in band 2 and 0.25 Samples in band 3. Together the Samples across Bands represent one Pixel. The Pixel Value Range is then computed by multiplying the Average Number of Samples per Pixel by the Sample Value Range.

Table 4-1: Pixel Value Range for Various Color Sampling Formats

Color Sampling Format	3-Band color (Average Samples/Band)			Average Samples/ Pixel	Sample Value Range (bits/Sample)		
	Band 1	Band 2	Band 3		8	10	12
					Pixel Value Range (bits/Pixel)		
4:4:4	1	1	1	3	24	30	36
4:2:2	1	0.5	0.5	2	16	20	24
4:2:0	1	0.25	0.25	1.5	12	15	18

Chapter 5

Dissemination

Motion Imagery Data is often produced some distance away from where it is controlled and/or exploited. The action of transmitting Motion Imagery Data from a source (i.e. Imager, Platform or Control Station) to one or more users is called Dissemination. Transmitting Motion Imagery Data can affect end users in two ways: Quality and Latency. Motion Imagery quality is impacted by the compression applied to the Motion Imagery and data losses during transmission. Similarly, Metadata can also be impacted by data losses.

Latency is a measure of amount of the time it takes to move data from one point to another in a Motion Imagery System. Latency is impacted by the compression of the Motion Imagery and the transmission path taken. Total Latency is the elapsed time from an occurrence in the Scene to when that occurrence is viewed in the Motion Imagery at its destination. When Total Latency is significant, a platform pilot may not be able to accurately control the Imager(s), and an end user may not be able to coordinate with other users or Intel sources in real time. Therefore, minimizing Total Latency is an overarching design goal, especially for systems used for real time applications. There is always a balance between Quality and Latency as both are difficult to optimize at one time.

While the subject of transmission can be extensive, in this section common methods endorsed by the MISP for Dissemination for Motion Imagery Data are discussed.

5.1 Background

Although the MISP does not levy requirements on the transmission of Motion Imagery Data, the MISP does levy requirements on the Quality of Motion Imagery, which can be greatly impacted by the transmission; understanding some basic methods for transmission is beneficial. The health of a delivery path from Imager through Exploitation depends on many factors, and begins with the method of transmission.

5.1.1 Transmission Methods

There are three transmission methods typically used in MISP applications: Wireless, Wired and Mixed Use.

5.1.1.1 Wireless

Wireless transmission generally assumes a radio link, such as from an airborne platform to a ground station. Although wireless technologies are designed to support varied applications and have different performance criteria, they are susceptible to interference from other communications signals. Interference introduces distortion into the transmitted signal, which can cause data errors. Because errors in wireless transmission are anticipated, methods to detect and repair errors often are provided; for example, Forward Error Correction is one popular method

used in a digital link. Such processes add additional overhead to the data transmitted, and they are limited to correcting certain types of errors.

5.1.1.2 Wired

Wired transmission can be divided into circuit-switched and packet-switched technologies. In circuit-switching a dedicated channel is established for the duration of the transmission; for example, a Serial Digital Interface (SDI) connection between a Sensor and an Encoder. Packet-switching, on the other hand, divides messages into packets and sends each packet individually with an accompanying destination address. Internet Protocol (IP) is based on packet-switching.

5.1.1.3 Mixed Use

In a network infrastructure, a mix of wireless and wired transmission methods is usually present. For example, Motion Imager Data from an airborne platform might be transmitted wirelessly to a satellite, relayed from the satellite to a ground receiver, and then transmitted over a wired IP network. Each method of transmission has its own susceptibility to errors that must be understood by developers when implementing a Motion Imagery System, and by users who receive and use the data.

5.1.1.4 Bandwidth

Wired transmission, in general, offers greater bandwidth capacity than wireless; this has important implications in the dissemination of Motion Imagery Data. Because of the large data characteristics of Motion Imagery, compression is needed when delivering Motion Imagery over the more bandwidth-constrained wireless link. Compression and subsequent encoding increases the complexity of the data, which makes it susceptible to errors introduced in transmission.

5.1.2 Internet Protocols

Internet Protocols represent a family of protocols used in an Internet packet-switching network to transmit data from one system to another. Table 5-1 provides information about the Internet Protocol family.

Table 5-1: Internet Protocols

Protocol Name	Description
Internet Protocol (IP)	The principle communications protocol for relaying packets across networks. IP data packets (datagrams) are sent from a transmitting to receiving system using switches and routers. IP is a low-level protocol that does not guarantee delivery, or when data arrives it will be correct (i.e. it could be corrupted).
User Data Protocol (UDP/IP)	UDP [1] uses a simple transport layer protocol based on IP. It does not guarantee data delivery or that data packets arrive in order. UDP specifies a network Port that enables multiple data sources from one system to be transmitted to multiple receiving systems. Data sent from one system to multiple systems is called multicasting. UDP provides one of the fastest methods of transmitting data to a receiver, which makes it suitable for time-sensitive applications (low latency). UDP multicasting is used in delivering Motion Imagery Data to multiple systems at once, which reduces overall network bandwidth.

Transmission Control Protocol (TCP/IP)	TCP [2] is a transport layer protocol that provides reliable guaranteed delivery of data. However, TCP does not guarantee time-sensitive delivery of data, but finds use in the transfer of non-time-sensitive data, such as Motion Imagery Data files.
--	---

When UDP/IP is used there are several types of packet errors that can occur as shown in Table 5-2. These errors can affect any protocol or data that uses UDP/IP (i.e. RTP).

Table 5-2: UDP Error Types

Error Type	Description
Packet Loss	Packets can be lost in a number of ways, such as network routers/switches being overwhelmed or network devices physically disconnected. When routers/switches are overwhelmed they will discard packets, which are then forever lost to all downstream devices. Other causes of packet loss include poor wiring and faulty equipment; these can cause intermittent packet loss and be hard to detect.
Packet Corrupted	Packets can be corrupted during the transmission from one device to another. Corruption can be caused by faulty equipment, poor wiring or from interference. Interference is primarily an issue with wireless technologies, although crosstalk in wired technologies can also be problematic. When routers/switches receive a packet and UDP error checking determines that the packet is corrupted, the packet is dropped and lost to a receiving system (see Packet Loss). If a corrupted packet passes its UDP error check, the corrupted packet could go undetected unless further error detection methods are used.
Packet Out of Order	Networks usually contain more than one router/switch, and typically there is more than one path for transmitting an IP packet from a source to a destination. Packets that take different paths may arrive at a destination out of the order they were transmitted. This condition is not detectable by UDP error checks, so other means for detecting and possibly reordering the packets need to come from additional information supplied within the transmitted data.

5.1.2.1 MPEG-2 TS Packets

The MPEG-2 Transport Stream (MPEG-2 TS [3]) is a widely used Container for disseminating Motion Imagery Data. For example, Motion Imagery Data transmitted from an airborne platform is typically in a MPEG-2 TS Container, as well as to points along a network that supports Exploitation. Developed originally for wireless transmission of television signals, MPEG-2 TS is organized as successive 188-byte data packets with each packet including a 4-bit continuity count. This count can be used to detect whether a packet is either missing or received out of order; however, because of the small size of the continuity counter it only detects a small percentage of the possible discontinuities. MPEG-2 TS is commonly used in delivering Motion Imagery Data over IP as well. The MISP has standardized how to insert MPEG-2 TS packets into UDP packets, see MISB ST 1402 [4] Table 5-3 describes the effects of UDP errors on the MPEG-2 TS.

Table 5-3: MPEG-2 TS Error Types

Error Type	Description
Packet Loss	<p>There are several types of packet loss in MPEG-2 TS. The first occurs when one (or more) UDP packet(s) are discarded. Up to seven MPEG-2 TS packets can be encapsulated into one UDP packet. Loss of one UDP packet can mean the loss of up to seven MPEG-2 TS packets. Such a loss can be detrimental to the decompression of the Motion Imagery, and the effects range from a Decoder that may stop working to intermittent losses of imagery. This significantly impacts Exploitation.</p> <p>A second type of packet loss is more localized to an individual MPEG-2 TS packet. Here, the internal information within a packet may be incorrect; this could result from a malfunctioning system component, or corruption in transmission. A packet could be discarded by receiving equipment if the error is seen as an ill-formed packet. Depending on the contents of a discarded packet the effect could be major (i.e. timing or important decode information) or minor (i.e. loss of a portion of the imagery).</p> <p>In both types of packet loss, when a packet contains Metadata the information is likely unrecoverable.</p>
Packet Corrupted	<p>A MPEG-2 TS packet may be corrupt resulting from a system issue, such as Encoder or Transport Stream multiplexer malfunction. Or information within a packet can become corrupted in transit. The packet may appear to be properly formed – and therefore not discarded, but the data contained inside is not meaningful. Issues like these are not easily diagnosed.</p>
Packet Out of Order	<p>As discussed in Table 5-2, out of order packets generally result from network device operation and varied network paths data may take. The 4-bit continuity count in each MPEG-2 TS packet provides a limited indication of packet sequence order; however, without information in advance on how many MPEG-2 TS packets are in a UDP packet it may be difficult to determine the actual MPEG-2 TS packet order.</p>

5.1.2.2 Real Time Protocol (RTP)

RTP [5] is designed for end-to-end, real-time transfer of data. RTP was specifically designed for delivery of A/V (Audio/Video) services over IP. Each data type (i.e. Motion Imagery, Metadata, and Audio) is delivered as an independent data stream. Relational timing information for synchronizing the individual data streams at a receiver is published in Real Time Control Protocol (RTCP) – a companion protocol. RTP is encapsulated in UDP/IP, and includes a timestamp for synchronization and sequence numbers that aide in packet loss and reordered packet detection. RTP/RTCP is typically considered for bandwidth-constrained environments, where a choice among the supported data types can be made.

There are also advantages to encapsulating a MPEG-2 TS into RTP; in fact, this method is widely used by the commercial industry in long-haul delivery of video over IP. A receiver can use the RTP timestamp to measure inter-packet jitter to estimate the stress occurring in a network. Such information also indicates potential Decoder buffer over/under flows, which could cause Decoder failure. The RTP sequence number is 16-bits; much larger than the MPEG-2 TS 4-bit packet count, which enables a wider detection range for lost and reordered packets.

Delivery of Motion Imagery Data using RTP is subject to the errors similarly found in MPEG-2 TS (see Table 5-3). In RTP, however, data is organized into larger packets, which can be as large as the limits (called the Maximum Transmission Unit, or MTU) of UDP for a particular medium. A lost packet of RTP may have a greater effect than a lost packet of MPEG-2 TS as it contains more data; again, the impact to decompression and Exploitation would depend on the data contained within the packet.

Appendix A References

- [1] IETF RFC 768 User Datagram Protocol, Aug 1980.
- [2] IETF RFC 793 Transmission Control Protocol, Sep 1981.
- [3] ISO/IEC 13818-1:2013 Information technology - Generic coding of moving pictures and associated audio information: Systems.
- [4] MISB ST 1402 MPEG-2 Transport of Compressed Motion Imagery and Metadata, Feb 2014.
- [5] IETF RFC 3550 RTP: A Transport Protocol for Real-Time Applications, Jul 2003.
- [6] MISB ST 0804.4 Real-Time Protocol for Motion Imagery and Metadata, Feb 2014.
- [7] MISP-2015.1 Motion Imagery Standards Profile, Oct 2014.

Appendix B Acronyms

FMV	Full Motion Video
IP	Internet Protocol
IR	Infrared
ISO	International Standards Organization
CLV	Key Length Value
MI	Motion Imagery
MISB	Motion Imagery Standards Board
MISP	Motion Imagery Standards Profile
MTU	Maximum Transmission Unit
RP	Recommended Practice
RTCP	Real Time Control Protocol
RTP	Real Time Protocol
SDI	Serial Digital Interface
SMPTE	Society of Motion Picture and Television Engineers
ST	Standard
TCP	Transmission Control Protocol
TS	MPEG-2 Transport Stream
UDP	User Data Protocol
UTC	Coordinated Universal Time