ISO/IEC JTC1/SC29/WG1 n101225
108th Meeting – Daejeon, Korea–30 June - 4 July 2025

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANIZATION INTERNATIONALE DE NORMALISATION

ISO/IEC JTC 1/SC 29/WG1
(ITU-T SG16)

Coding of Still Pictures

JBIG	**JPEG**
Joint Bi-level Image Experts Group	Joint Photographic Experts Group

TITLE: Use Cases and Requirements for JPEG GMIS v3.0

EDITORS: Reji Mathew (reji@unsw.edu.au)
Oh-Jin Kwon (ojkwon@sejong.ac.kr)

STATUS: Approved

REQUESTED ACTION: Distribution

DISTRIBUTION: Public

Contact:
ISO/IEC JTC 1/SC 29/WG1 Convener – Prof. Touradj Ebrahimi
EPFL/STI/IEL/GR-EB, Station 11, CH-1015 Lausanne, Switzerland
Tel: +41 21 693 2606, Fax: +41 21 693 7600, E-mail: Touradj.Ebrahimi@epfl.ch

Editorial Comments

This is a living document that goes through iterations. Proposals for revisions of the text can be delivered to the editor Reji Mathew and Oh-Jin Kwon, by downloading this document, editing it using track changes, and sending it to reji@unsw.edu.au or ojkwon@sejong.ac.kr.

If you have interest in JPEG GMIS, please subscribe to the email reflector, via the following link: http://listregistration.jpeg.org.

Use Cases and Requirements for JPEG GMIS v3.0

1 Introduction

The number of digital photos on personal devices like mobiles and the internet is increasing rapidly. More than billions of images are generated a day, transmitted, and stored in social networking environments. It is noticed that the current trend of taking pictures does not generate a single image anymore but it often generates a set of similar multiple images at a time and place. Figure 1 shows examples of these multiple images.

Figure 1 - Examples of similar multiple images

These images are mostly redundant in their colors, shapes, patterns, etc., and can be transmitted/stored together as a set of images instead of being transmitted/stored individually. It is desirable to have a way to store and transmit a set of related images using a single file object that contains the whole set. There are currently four main approaches that are used in practice:

1. Use a format concatenating multiple images into a file like Multi-Picture Format (extension .mpo). Various devices including digital cameras, smartphones, camcorders, and game consoles, use it to store 3D images, preview images, stereoscopic images, etc. The disadvantages of this method are that there is a risk of misinterpretation of this concatenation as a single image and any modern compression technique utilizing the inter-image redundancies is not employed for further compression.
2. Treat the image set as an animation, i.e. as an ordered and timed sequence, and store it using an animated image format like GIF, MJPEG, animated WebP, AVIF, HEIF, JPEG 2000, JPEG XL, or a video format. This method is not very suitable since it forces a specific playback order and timing, and it is not convenient to add/remove images from a set or to extract single images from a set. It also assumes that the images have identical dimensions; if the set e.g. contains a mix of portrait and landscape aspect ratios, padding or cropping will have to be performed. Furthermore, it normally utilizes inter-frame pixel redundancies for compressing multiple images

while similar multiple images shown in Figure 1 show different redundancies that are not suitable for existing video/animation coding tools.

3. Treat the image set as multiple pages and embed them in a multi-page format like PDF, TIFF, or even Microsoft Word. This method is also not ideal since it forces a specific order and page/canvas layout, and it implies the 'overkill' of using complicated and/or proprietary formats that can also do many other things.
4. Use general-purpose archiving tools like ZIP, RAR, 7z, tar, etc to store the set of images. The 'Comic Book Archive' de-facto standard (.cbz, .cbr, etc) works this way. A disadvantage of this approach is that inter-image redundancies can only be exploited to the extent allowed by general-purpose compression; there are no image-specific inter-image coding tools.

None of these existing approaches provide a satisfactory solution to obtain a lightweight yet effective container format for multiple images that allows interactive and non-timed navigation.

On the other hand, new JPEG standards such as JLINK (ISO/IEC 19566-7) and JPEG Snack (ISO/IEC 19566-8) have been established to handle multiple images. They may generate the 'mother' JFIF file which may embed other 'child' images independently coded using JUMBF (ISO/IEC 19566-5) metadata as exemplified in Figure 2 and this trend will cause larger-sized mother JFIF files so that the industrial demand of compressing these mother JFIF files with higher compression will increase rapidly.

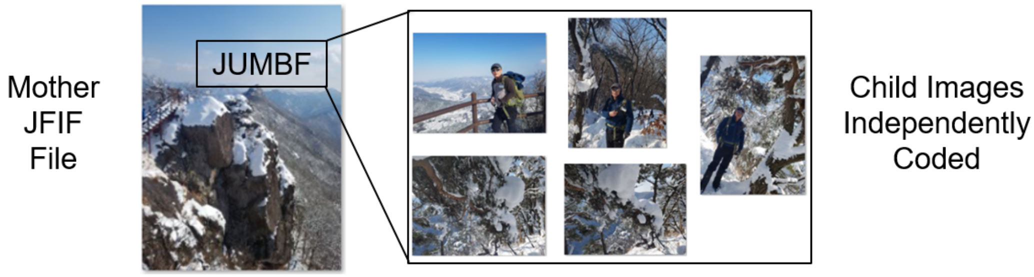


Figure 2 - Example of mother JFIF file embedding multiple child images independently coded

Compared to the aforementioned existing approaches to binding multiple images into a single file, one of the merits of this mother JFIF file is that it does not need any unbinding process to show the mother JFIF image which is normally representative of all multiple images involved. Unfortunately, JLINK and JPEG Snack only allow child images independently coded. We cannot employ any kind of inter-image coding methods for reducing the size of child images. The JPEG GMIS intends to compress child images by allowing mother JPEG files (or any other JUMBF-capable format) to be able to embed child images coded by exploiting some kinds of inter-image redundancy as shown in Figure 3. A decoder perspective of the JPEG GMIS framework is shown in Figure 4.

Figure 3 - Basic structure of mother JPEG file using JPEG GMIS

2 Scope

The scope of JPEG GMIS is the creation of a standard to efficiently represent sets of related images and the relationships between these images to facilitate composition, navigation, display, processing, storage and transmission of image sets.

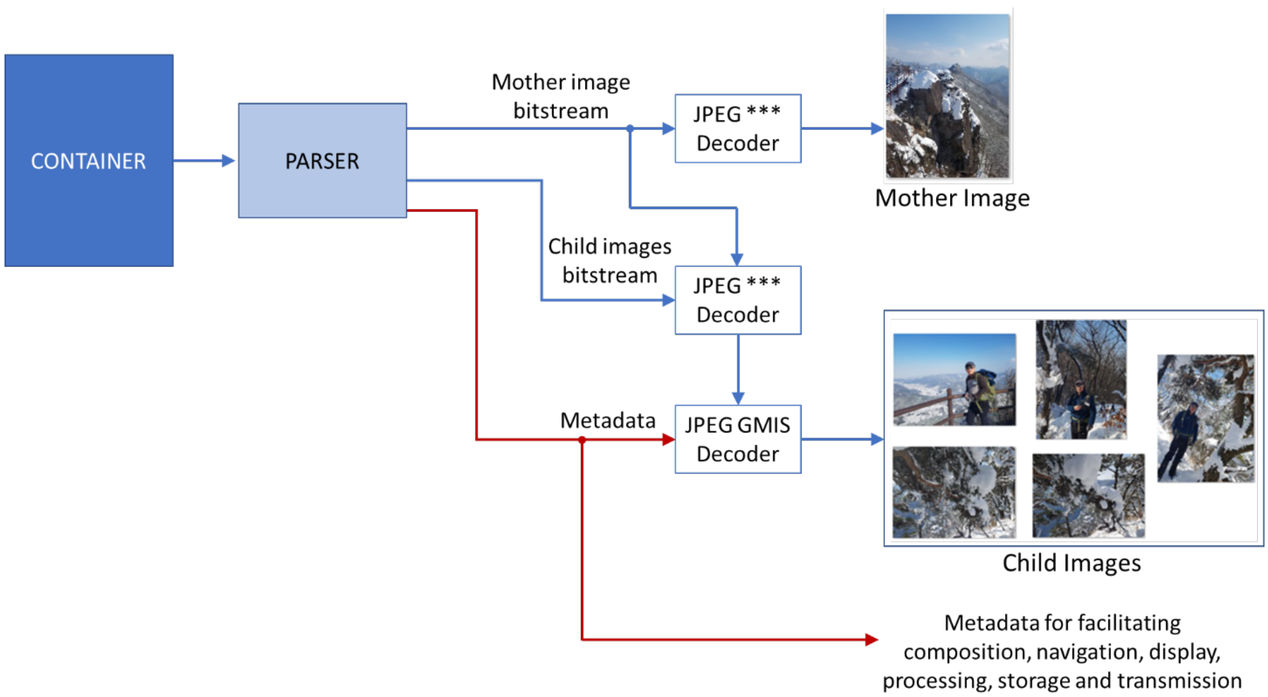


Figure 4 - JPEG GMIS decoding framework, where JPEG *** Decoder refers to a JPEG standardised decoder

A JPEG GMIS representation consists of multiple compressed image bitstreams with associated metadata. As shown in Figure 4, all image bitstreams are required to conform to an existing JPEG coding standard. The metadata that accompany stored images describe relationships between these images and is critical to the features and functionalities that can be provided by JPEG GMIS.

The contributions of the metadata can be classified into two broad categories. First, the metadata information can facilitate informed navigation of the image set, enabling features such as interactive searching and potentially composition, display, processing, and transmission based on user preferences. Second, the inter-image relationships conveyed by the metadata can be used to exploit redundancy between images to improve compression efficiency.

Existing standardised image container file formats in general are limited to providing a scheduled navigation or a predefined composition of the image set. This includes, for example, a user or application explicitly specifying the images to be retrieved and the order in which to render these decoded images. In addition to this basic feature, JPEG GMIS seeks to utilise the relationship between images to enable informed navigations. This includes discovering the most appropriate images to retrieve and determining a suitable rendering order, in response to a user query or preference based on inter-image relationships.

The relationships between images that need to be conveyed by the metadata will vary depending on the target application and type of image set. JPEG GMIS is extensible allowing new types of inter-image relationships and their metadata to be standardised, where the availability of such new data provides improvements in either the efficient navigation of image sets or compression performance.

This document describes a variety of use cases of JPEG GMIS. Included in these discussions are examples of metadata and its utility for the various image types and applications considered.

3 Use cases

This section presents several, but not necessarily all, use cases relevant to the JPEG GMIS standard.

3.1 Sharing a set of images: main use case

This subsection includes use cases related to sharing multiple images with others on social networks.

3.1.1 Event shots

Andy likes sharing his photos with friends through social networking services (SNS). He compressed and uploaded a series of photos that he took when he climbed the Rocky Mountains last week. His friends enjoy the photos by accessing the SNS site.

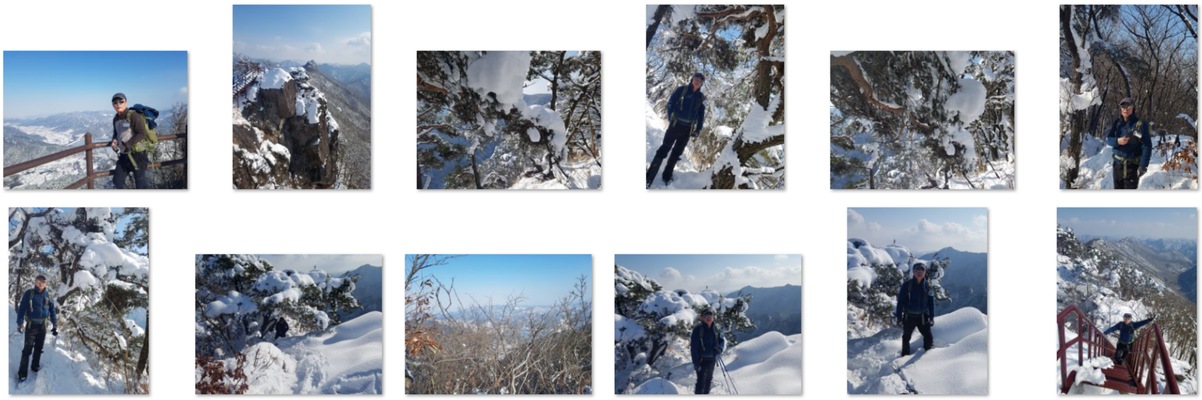


Figure 5 - Example of event shots belonging to one category

Andy's friends notice that Andy's photos are serviced in the form of categorized photo albums. Each category consists of a set of images with directories, showing a representative JPEG image and related information such as the time and place information that Andy took the photos. An example of an image set belonging one specific category is shown in Figure 5. Andy's friends enjoy Andy's album by navigating each category. They may skip some categories with a look at only the representative image or the thumbnail images for the categories or click one of the images in a category for looking at the full-sized image. Examples of image sets and corresponding representative images for various categories are shown in Figure 6.

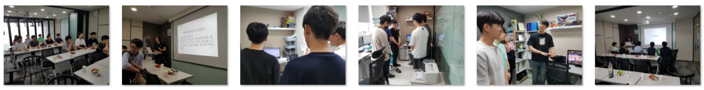
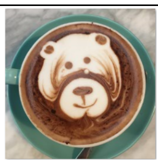
Category	Representative Image	Other Images
1		
2		
3		
4		

Figure 6 - Example of categorized photo album

Andy's album also includes sets of photos taken by his new smartphone in which multiple cameras are installed. Multiple zoomed, PIP, and split photos with information about the relationship of photos are provided by this smartphone. Multiple views of a scene, captured with different zoom factors and camera orientations are shown in Figure 7.

Figure 7 - Example of multiple images of a scene, related by camera zoom and viewing angle

In this use case scenario, metadata that describes geometrical relationships between images can be beneficial for compression.

3.1.2 Reporting a set of images

Dorothy had a car accident on her way home from work. As shown in Figure 8 she took several photos of the damaged car and reported them to the insurance company in the form of one compressed file.

Figure 8 - Multiple images of a single object of interest required to be associated and stored together

John, a plumber, takes a set of several pictures of the water pipe to be repaired before starting work. He also takes a set of pictures of the repaired water pipe and posts the situation before and after the repair on his blog using the JPEG file format to advertise his work; refer to Figure 9 for an example set of related images. Sometimes, he adds or removes such JPEG files on his blog.

Figure 9: Multiple images of a scene required to be stored together and rendered in a specific viewing order

Metadata in this case can be useful for defining a viewing order of images (e.g., image before repair viewed first and then images after repair). Furthermore, metadata in the form of geometrical relationships can also be useful for compression if there is sufficient overlap between images in the set.

3.1.3 Continuous shots

Speed skating is Peter's favorite sport. He took a speed skater's finish using his high-speed camera application and obtained a series of images. He selected, cropped, enlarged, and touched some of them. An example of a set of related images is shown in Figure 10. He shares a JPEG file which is the compressed form of this set of images with his friends through an SNS site. A representative image of this file with a mark is shown to his friends. When his friends click the mark, all the images are displayed. This file is saved in short-term storage of the SNS server for a certain period. When this file is not used for long, it is further encoded by a method that achieves much higher compression but demands a longer decoding time and moves to long-term economic storage.

Figure 10 - Example of a set of continuous shots comprising of images at different resolutions and aspect ratios

For this use case, metadata that is capable of describing geometric relationships between images of differing resolution, orientation and aspect ratio can be useful for efficient compression, allowing predictive coding to be performed between images of different size.

3.2 Multiple images for compositing higher quality image

This subsection includes use cases in which multiple images are used for compositing a higher quality and/or more informative image. While accompanying JPEG GMIS metadata can facilitate the compositing of a new image, the actual compositing process is outside the scope of this standard.

3.2.1 Multi-focus images

Hannah is a professional photographer. She is frustrated when she tries to capture a picture of several objects at different distances. When the foreground object is in focus, the background is out of focus, and vice versa. She has a solution to this problem. She captures images at various focuses and sends them to her assistant who operates a computer program to fuse them into an all-in-focus image, where all visible objects are in focus. An example of a focus stack and a corresponding generated fused image is shown in Figure 11,

Fused Image	Multi-Focus Images
	

Figure 11 - Example of image set comprised of a single fused all-in-focus image and multiple component images focused at different regions of the scene

3.2.2 Multi-exposure images

Hannah also takes a picture that comprises shadows or highlighted areas, she is faced with the challenge of setting the appropriate exposure. Fortunately, her assistant may use a high-dynamic-range (HDR) program to resolve this problem by fusing multi-exposed images of the same scene. For fast processing, the HDR program provides reduced-size overviewing images for the assistant to select the images involved in the fusion. The assistant may also select the fusion algorithm and generate different types of fused images with information on how to fuse them. An exposure stack along with a generated HDR fused image is shown in Figure 12.

Fused Image	Multi-Exposure Images
	

Figure 12 – Example of image set comprised of a single fused HDR image and multiple component images captured with different exposure times

3.2.3 Multiple images for compositing panorama and 360° image

Jennifer is a tour guide. She took several images to promote new places. She generated panorama and/or 360° images using a stitching program, saved the original and the stitched image files with the information on how to stitch them in the form of a set of JPEG files, and posted the stitched images on her website. An example case is shown in Figure 13. After a few days, she found discontinuities on some stitched images due to her carelessness in setting the input parameters of the stitching program. She downloads the images, finds the wrong parameters, runs the program with proper parameters again, succeeds in generating the stitched images without any discontinuity, and reposts them.

Stitched Image

Multiple Images for Stitching

Figure 13 - Example of image set comprised of a single panorama image and multiple component images captured from different viewing angles

Tom, who works for a global company, often hosts online meetings with overseas branches. Sometimes, he wanted to see all the other attendants in the conference room at a specific time during the meeting but he couldn't because of the narrow view of the camera. To solve this problem, he recently uses an application that captures all the people in the conference room using multiple cameras or PTZ cameras and provides a panorama image covering all the attendants at a time. An example of a composited wide angle view image, along with a number of captured narrow view images are shown in Figure 14.

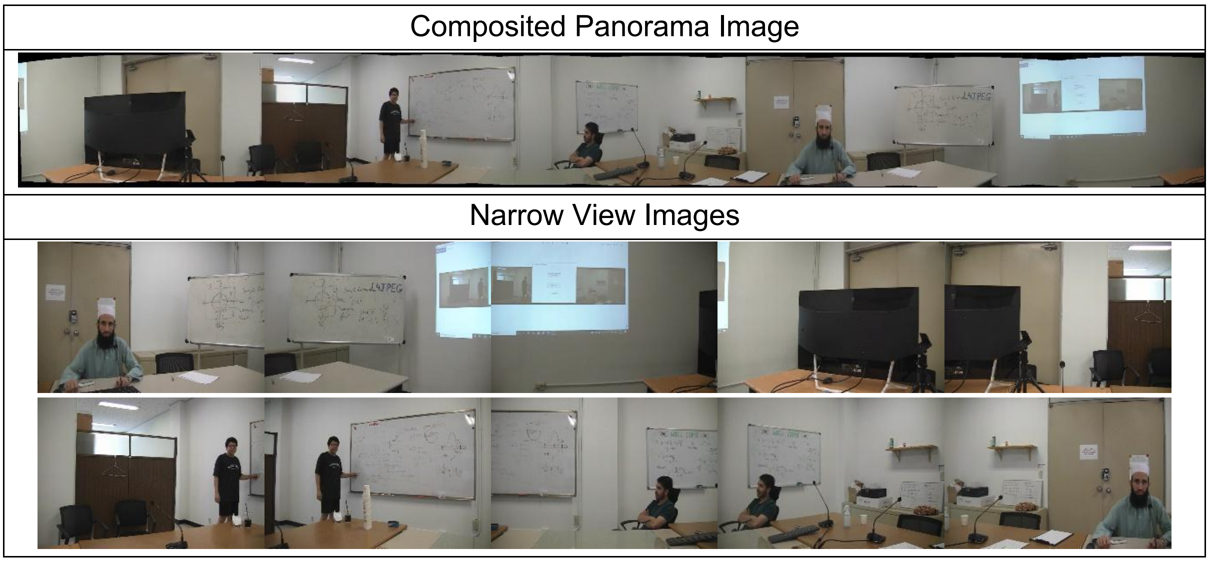
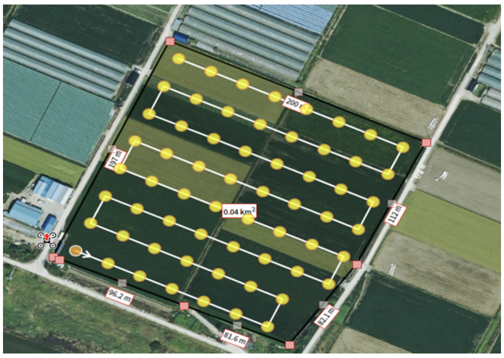


Figure 14 - Example of image set comprised of a single 360° view image and multiple component narrow view images

3.2.4 Multiple images for compositing feature map

Neighbor System is a company that specializes in feature map services. One notable service they offer is the provision of feature maps that showcase the real-time growth status of crops. These feature maps are generated by capturing aerial photos, stitching them into a single large photo that encompasses the entire service area, and superimposing it onto a general overview map; refer to Figure 15. To enhance the map's display efficiency, the stitched photo is subdivided into smaller-sized images, which are transmitted and seamlessly integrated onto the general overview map.

Scan Path for Capturing Aerial Photos	Captured Aerial Photos
	
Stitched Image	Smaller-sized Images for Display Efficiency
	

Displayed Feature Map Superimposed onto a General Overview Map

Figure 15 - Examples of drone images: captured aerial photos are used to form a larger stitched image which is then sub-divided to smaller images for displaying local regions superimposed with overview maps.

These use cases demonstrate different relationships between images in a set. Metadata describing inter-image relationships can be useful in facilitating interactive searching of image sets. Efficiently navigating through an image set to identify specific images or regions of interest becomes increasingly important for larger sets where manual searching is not feasible. For focal stacks, metadata in the form of an auxiliary image such as a depth map with accompanying camera parameter details, can enable interactive searching to locate individual images for which a user defined point or region is in focus. For image sets that describe a scene captured from multiple viewpoints, camera pose details corresponding to each image can enable interactive searching for images from the set that approximate a user defined viewing angle and location.

Metadata that can adequately describe inter-image relationships can also be used for improving compression of the image sets considered above. Descriptions of image blurring or sharpening filters can be useful metadata for exploiting redundancy between images of a multi-focus image set. Similarly, metadata describing pixel value scale and offset factors can be used for predictive coding of a multi-exposure image set. Geometric relationships can be utilised for coding image sets used for image stitching applications including panoramic or wide angle view generation.

A GMIS aware postprocessing system can potentially use the stored metadata to assist in forming composited images from individual images in the coded set. As an example, for focal stacks, accompanying metadata which can locate in focus pixel locations, can also be utilised to composite an all-in-focus image. In this way, accompanying metadata can be useful for both compression and postprocessing rendering applications.

3.3 Multi-spectral images

Mark is an employee of the meteorological administration. He utilizes multiple images of various spectra provided by modern weather satellites to make accurate weather forecasts; refer to Figure 16. The images are in the form of a set of images including the metadata relating to these images. Sometimes, this set of images is updated by modifying the included images and the metadata.

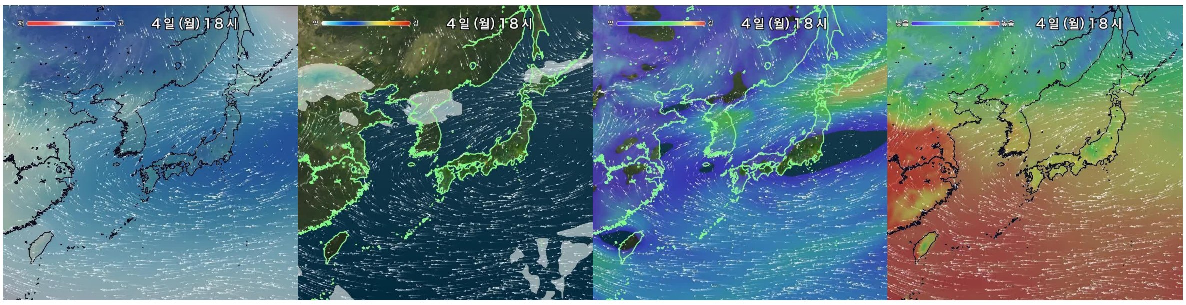


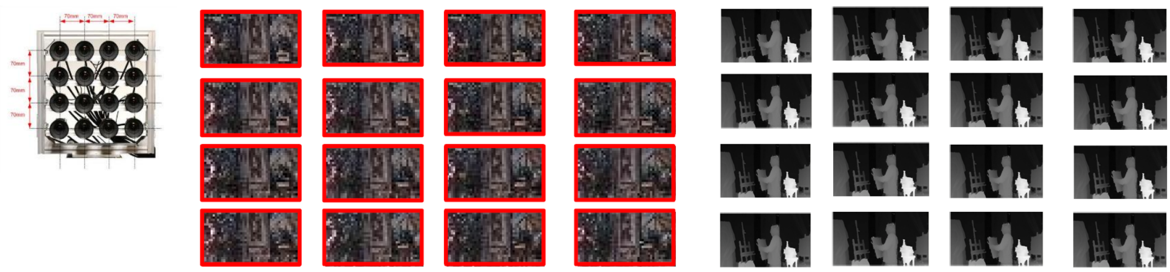
Figure 16 - Example of image set comprised of multi-spectral satellite imagery

3.4 Multiple images for compositing a 3D image

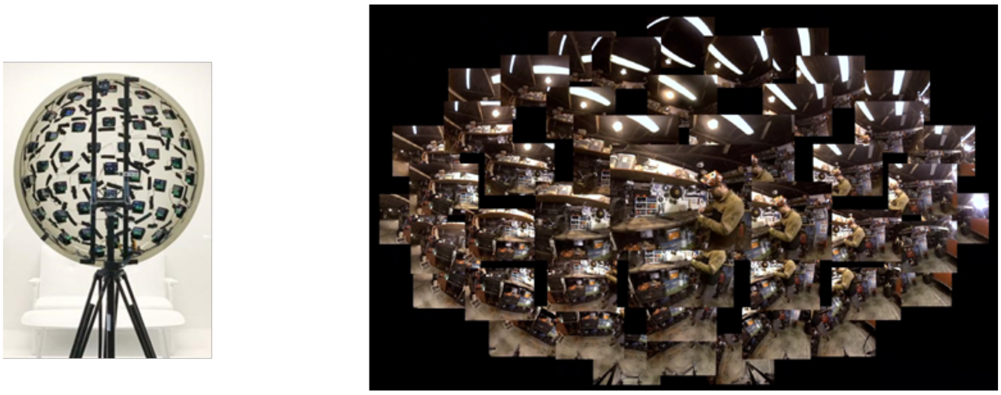
Jane is an artist for a company providing order-based 3D images for various customers. She first composites 3D images from various types of multiple images such as stereoscopic, multi-plane, multi-view, multi-depth, and multi-sphere images as shown in **Figure** 17. She then touches them artistically by coloring, texturing, and lightening. She saves the input multiple images and the information about her composite as a JPEG image file. Sometimes, the customer reorders her art with different composites and touches. Jane retrieves previous multiple images, re-composites, retouches, and provides her new art to the customer.

(a) Stereoscopic Image Pair

(b) Multi-Plane Images



(c) Multi-View and Depth Images:



(d) Multi-Sphere Images

Figure 17 - Example of image sets for 3D view composition

Metadata can facilitate view warping and can potentially provide substantial improvement to coding of multi-view images. Such metadata can also be useful for searching and navigating content based on geometric relationships between images. For example, the task of identifying images in a set that have considerable overlap with a current chosen image, can be aided by accompanying metadata.

3.5 Multi-modality medical images

Hospitals use several medical imaging modalities to accurately determine patients' diseases. Unlike ordinary digital cameras, each medical modality generates a large number of images with similar characteristics depending on the modality. The generated medical images are stored or fused for a display to help the treatment and diagnosis of the patients and sometimes transferred to the patients through the network or the portable storage when they are requested by the patient. A set of multimode medical images and generated image fusions are shown in Figure 18.

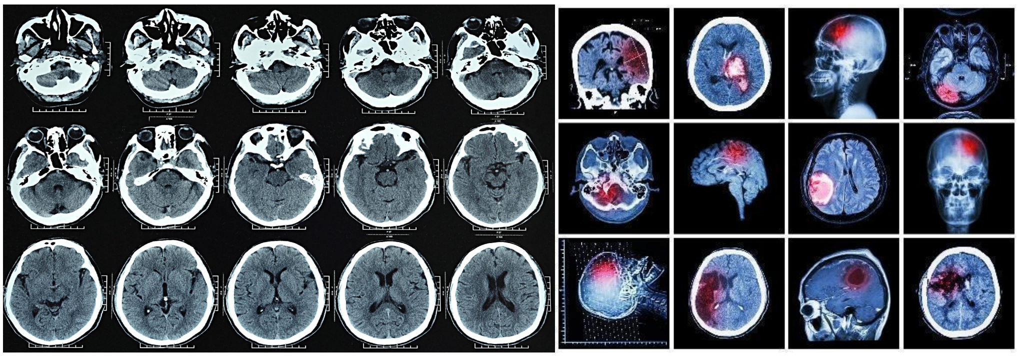


Figure 18 - Example of an image set comprised of multimode medical images (left), used by image fusion algorithms to generate new imagery for display and analysis (right)

3.6 Holo-tomography Imaging

Holo-tomography imaging technology enables 3D visualisation of living cells and tissues without labeling or staining, offering label-free, quantitative imaging at high resolution. It combines the principles of holography and tomography, allowing for the reconstruction of 3D refractive index distributions. Currently, the computed 3D volumetric data is stored as a set of independently coded 2D images with each image corresponding to a slice along the axial direction. Metadata that can describe relationships between individual slices can potentially enable joint coding of these slices, exploiting redundancy that may be present across slices in the computed volumetric representation.

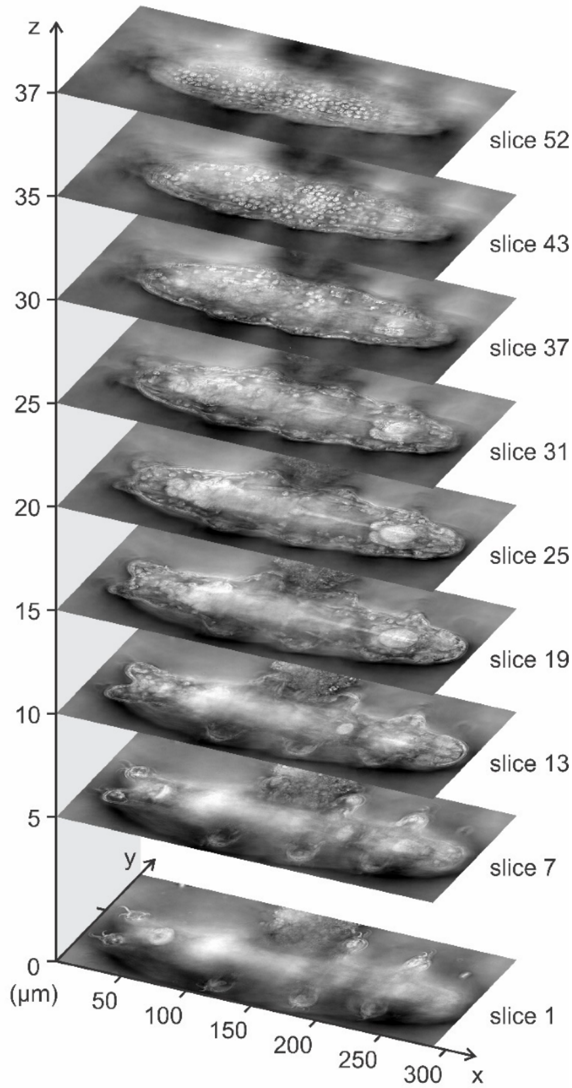


Figure 19: Representative slices from a reconstructed holo-tomography stack[1]

3.7 Fourier Ptychographic Microscopy Imaging

[1] Hong, Minh-Triet, Giyoung Lee, and Young-Tae Chang. 2025. "A Non-Invasive, Label-Free Method for Examining Tardigrade Anatomy Using Holotomography" Tomography 11, no. 3: 34. https://doi.org/10.3390/tomography11030034

Fourier Ptychographic Microscopy (FPM) is a computational microscopy technique that can enhance the resolution and field of view of a conventional optical microscope. FPM relies on stitching together multiple, low resolution intensity images corresponding to different angles of incidence of an illuminating coherent light source. These low resolution images are then combined using an iterative phase retrieval algorithm to produce a final high resolution output image. In current applications, the set of acquired FPM intensity images are stored as independently coded images. Metadata describing relationships between these images can provide multiple benefits, including the potential of improving compression by exploiting redundancy that may be present among images in the set.

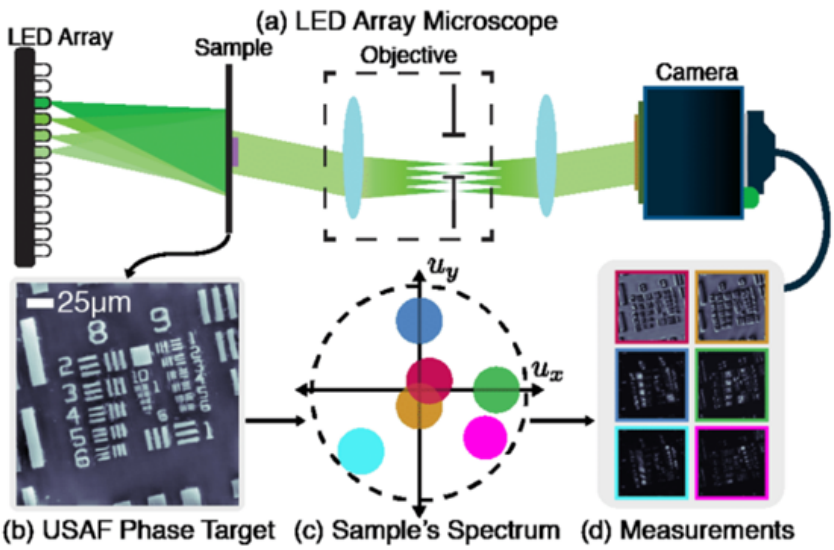
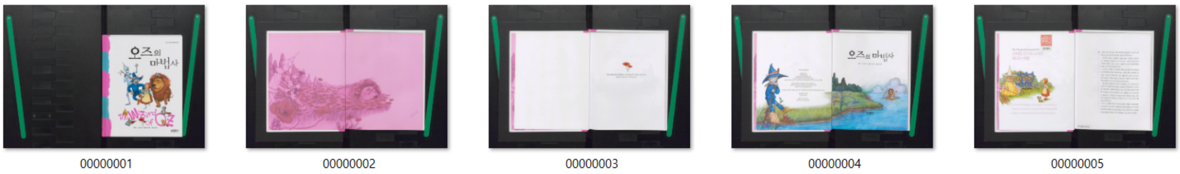


Figure 20: Fourier Ptychographic Microscopy (FPM) with an LED array source.[2]

3.8 Paging images

A company providing online service of old books collects the books expected to attract their customers after making contracts with the copyright owners for online servicing. The company scans the books by dedicated high-speed scanners, aligns the scanned pages, categorizes them, groups multiple pages into sets, and provides overviews and indexes for the sets to the customer. The books include various types: cartoons, drawings, graphics, and photos in addition to texts. Example scans are shown in Figure 21.

[2] Data-Driven Design for Fourier Ptychographic Microscopy, 2019 IEEE International Conference on Computational Photography (ICCP), DOI:10.1109/ICCPHOT.2019.8747339



(a) Scanned images containing text and drawings

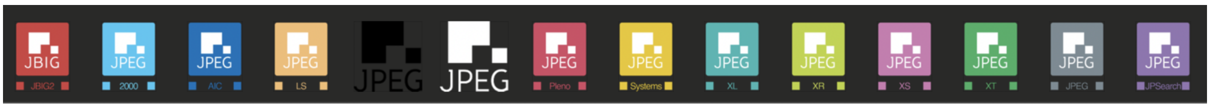
(b) Scanned images containing cartoon illustrations

Figure 21 - Examples of image sets comprised of scanned images from books

3.9 Multiple images with shared elements

3.9.1 Brand assets

Brand logos and other assets, such as those shown in Figure 22, are often created in a variety of color schemes, aspect ratios, levels of detail, etc. While for a specific use, only one image will be selected, the different variants nevertheless have shared elements that could be exploited to have a compact representation that can be used to share the asset collection.



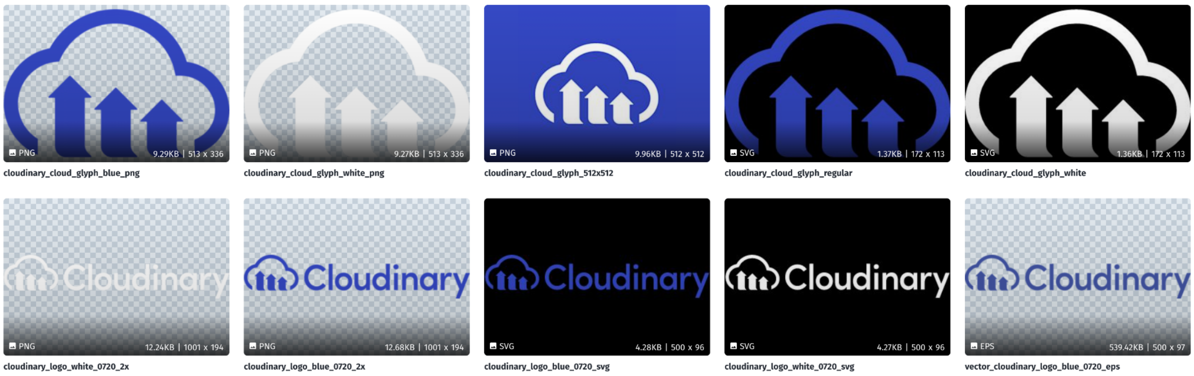


Figure 22 - Graphic image sets with shared elements

3.9.2 Product images with shared elements

Product images, as shown in Figure 23, are often processed to obtain multiple variants of the same original image, with various (or no) backgrounds, different overlays, effects, etc.

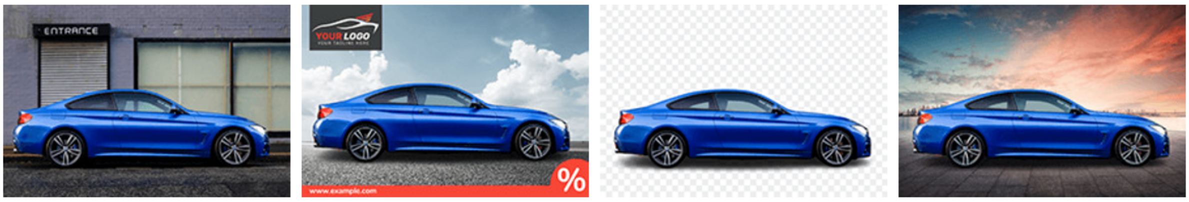


Figure 23 - Image set containing a shared product object with varying backgrounds and overlays

3.9.3 Multiple crops from the same source image

From a single large image, multiple images can be derived by cropping, to obtain different aspect ratios or different zoom levels, or to focus on different regions of interest. An example is shown in Figure 24.

Figure 24 - Image set comprised of a large source image and multiple crops at different zoom levels and aspect ratios

4 Requirements

JPEG GMIS considers various components that have their own requirements. Essential "core requirements" are identified with a 'shall' and "complementary requirements" are identified with a 'should'. The following requirements are relevant for JPEG GMIS:

4.1 Metadata requirements

A. JPEG GMIS shall support providing information about an image set which shall be stored together with the image codestreams in a file, especially information describing images and the relationship between images. These relationships can be temporal, spatial, spectral, or others.

B. JPEG GMIS shall be capable of providing metadata describing images and relationships between images, for a wide range of image acquisition configurations, including imagery captured by multiple non-identical cameras, at a plurality of arbitrary locations and orientations.

C. JPEG GMIS shall support metadata for a set of images providing information about a mother image of the set.

D. JPEG GMIS should support metadata for identifying overview thumbnails of images in a set.

E. JPEG GMIS should support metadata providing presentation instructions for viewing applications, e.g. a preferred navigation interface (clickable thumbnails, forward/back buttons, vertically scrolling document view, etc).

F. JPEG GMIS shall support metadata that allows synthesis of images from stored codestreams.

G. JPEG GMIS shall support metadata describing relationships between synthesised and decoded images.

4.2 Systems and Integrations requirements

A. JPEG GMIS shall support ISO box-based media format.

B. JPEG GMIS shall, for any chosen image from file, facilitate efficient discovery of all metadata that describe relationships among the chosen image and other images defined by the file.

C. JPEG GMIS shall support easy access to metadata describing relationships between images in a way that facilitates discovery of related image regions and determination of steps required to recover imagery of interest from the codestreams found within the file.

D. JPEG GMIS shall support the addition of new codestreams and relationship metadata without impacting the decoding or synthesis of existing images contained in file.

E. JPEG GMIS shall enable editing of stored images and relationship metadata by efficiently identifying all other metadata descriptions and codestreams that could be impacted and may need to be reevaluated or replaced.

F. JPEG GMIS shall support existing JPEG container file formats, such as JPEG XT Part 3 (JFIF plus box extensions) and JPEG 2000 Part 2 (ISO box-based media format) such that JPEG GMIS content can be mapped to existing file formats using boxes, structures and extension mechanisms as defined by corresponding JPEG container file formats.

G. JPEG GMIS shall be capable of mapping to a file, both codestreams and metadata from existing JPEG container file formats such as JPEG XT Part 3 and JPEG 2000 Part 2.

H. JPEG GMIS shall support existing JPEG Systems standards framework (JUMBF-based format) for providing proprietary or vendor specific information that describe relationships between images .

4.3 Coding requirements

A. Under the right conditions, JPEG GMIS shall allow the efficient coding of one set of related images into one file from which all the images in the set can be decoded.

B. JPEG GMIS shall provide one mother image, decodable with a non-GMIS-aware decoder compliant to an applicable existing JPEG coding standard.
C. JPEG GMIS shall support means to signal the decoding tools required to recover all other images.
D. JPEG GMIS shall support the encoding of all images in the set using one or more codestreams that conform to existing JPEG coding standards.
E. JPEG GMIS should support means to exploit the redundancy between images in the set that do not require the underlying codestream decoding procedures to be altered from those described in existing JPEG coding standards.
F. JPEG GMIS shall support means for efficiently accessing images in a set of images without completely decoding all the images in the set.
G. JPEG GMIS shall support flexible means to trade-off coding efficiency against random accessibility.
H. JPEG GMIS should allow easy adding and removing of images in a set of images.
I. JPEG GMIS should support a variety of image types: natural, medical, projected, graphical, 360-degree, multi-spectral, HDR, etc.
J. JPEG GMIS shall support coding sets of images with different resolutions, aspect ratios and orientations.
K. JPEG GMIS shall support coding sets of images that include both color and greyscale images.
L. JPEG GMIS should support low complexity implementation in terms of computation, memory, and power consumption.