

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

Coding of Still Pictures

JBIG

Joint Bi-level Image
Experts Group

JPEG

Joint Photographic
Experts Group

TITLE: JPEG 360 Metadata Use Cases

SOURCE: WG1

PROJECT: ISO/IEC 19566-6 (JPEG Systems Part 6)

STATUS: Final

REQUESTED ACTION: Publish

DISTRIBUTION: Public

Contact:

ISO/IEC JTC 1/SC 29/WG 1 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

1. Purpose of This Document

This document describes a variety of use cases for 360 imaging (defined below) which allow us to create requirements for the metadata that is included within the image file needed to support these use cases. It is expected that use cases will evolve over time, and that requirements will be staged over time.

2. Definition of 360 Imaging

In the context of this document, “360 imaging” (aka “omnidirectional imaging”) is defined similar to the Wikipedia definition below.

360 videos, also known as immersive videos[1] or spherical videos,[2] are video recordings where a view in every direction is recorded at the same time, shot using an omnidirectional camera or a collection of cameras. During playback the viewer has control of the viewing direction like a panorama. (https://en.wikipedia.org/wiki/360_video)

JPEG 360 extends this definition: a 360 image field-of-view ranging from a full sphere (i.e., the solid angle subtending 4π steradian) to spherical surfaces subtending a fraction of 4π steradian. Since the 360 image is constrained to the a spherical surface, there are only 3 degrees-of-freedom (3 DoF) corresponding to rotations (yaw, pitch, roll) around the three spatial axes. By convention, the center of the sphere is also the origin of the three spatial axes; however, the definition of mapping of the image projection wrt rotation about those axes are required to be provided with the 360 metadata. E.g, in an equirectangular projection, the yaw/pitch/roll of the corner pixels must be provided.

3. JPEG 360 Use Cases

In the subsections that follow, brief use cases from JPEG experts are captured. The use cases bring out certain interactions with 360 images that are considered relevant and interesting, but without being prescriptive with respect to implementation.

3.1. Action Sports Enthusiast

François uses 360 degrees panorama camera to record his paragliding adventures. His flight equipment (altimeter, barometer, compass, GPS, etc.) are streamed to the camera for capture. The 6DOF sensor for the camera rig is also streamed for capture. And the biometric sensors for the pilot are streamed to the camera for capture. Let’s assume the camera is set to capture a high resolution 360 degrees panorama image every 1 second, along with the aggregated sensor data.

When François returns to ground, the camera data is uploaded to a personal computer and/or cloud service. The basic sensor data can be used to link an image to a position on a map. François, however, subscribed to an enhanced service, running on his personal computer which able to determine which landmarks are visible in an image and augment with image with additional labels; e.g., Mont Blanc. The derived metadata components (Mont Blanc’s position in the image, and the position of the label in the image) are new metadata that are to be added to the François’ images to create enhanced interactions. François’ children enjoy learning geography when they see the images.

3.2. Engineering Smart Assistant

Tom is an engineer in charge of structural engineering for reinforced concrete. As there are many aged reinforced concrete constructions to be regularly inspected, he really feels that it is necessary to raise the efficiency of inspection. Tom has been storing a huge number of sample pictures regarding deterioration of concrete constructions. As a new technology trend, Tom has heard that advances in image recognition technology will become practical on mobile devices. He decided to take in the new technology to improve its effectiveness. Tom took pictures with his smart

phone, being careful to capture them with high fidelity to improve the analyses. Below is a sample picture applied for a dam, but the images and analyses are similar for tunnels and bridges.

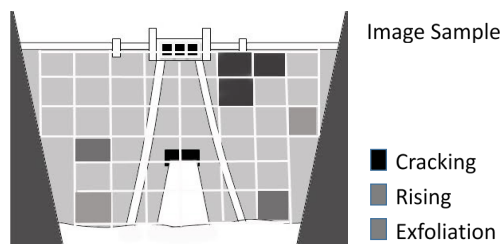


Figure 3.2 Engineering Smart Assistant

3.3. Memento Photo Collage

Alice has returned to Sydney after being a bridesmaid at her best friend’s destination-wedding in Cusco, Peru. Alice had a great time at the wedding, and really enjoyed a self-guided tour of the city using a smart guide developed by a Cusco native. Alice added a few days to visit Machu Pichu, and added a few days in Lima to explore the culture, food, nightlife, and beaches. Alice is very taken with the traditional Peruvian music, and also really liked a couple popular songs on the radio, so she downloaded those for her music collection.

On the long flight back to Sydney, Alice went through her photos and created a collage of photos and used the newly acquired music as the collage soundtrack that would play when Alice played the collage. The application she used to create the collage also allowed Alice to link to the full resolution image version of each of the photo tiles in her collage. The entire memento photo collage, including high resolution photos and soundtrack, was bundled into a single JPEG file so that it could be easily shared to her electronic photo frame and share the memento with friends; everyone could at least see the photo collage, and those with newer photo frames got to hear the soundtrack and enjoy the full-screen versions of the photos.



Figure 3.3 Memento Photo Collage

3.4. Navigating in Omnidirectional Images

Heidi captures a 360 picture of her grandfather’s house and garden. Later, when she reviews and edits the image, Heidi selects a region of interest (ROI) within the omnidirectional photo which can be viewed using a legacy JPEG viewer. Heidi also adds the default position the viewer should see when opening the 360 image. She sends this composite image to her friend Klara as an email attachment.

Klara receives Heidi’s email. Klara is able to view Heidi’s the ROI portion of photograph with a conventional JPEG viewer on her PC. However, her PC’s photo viewer application supports the JPEG 360 photos, so when she uses that application, Klara can choose to view either the ROI image or the 360 image. Klara’s phone also has a photo viewer application that supports JPEG 360 photos. Klara looks at Heidi’s photo with using her smartphone., and then decides

to switch to her PC to use her head mounted display (HMD) to have a more immersive and enjoyable experience which brings back memories of her life in the Alps. Klara decided that in the future, she wants to open the 360 image to a different default view, and she edits this property of the image. If she wants, Klara could change the ROI image by having the 360 photo editor replace it with her preference.

Heidi emails the URL of the omnidirectional photo to her childhood friend Peter, so Peter can access it with a 360 web-viewer using a browser on his computer. He can interact with the picture naturally without any noticeable delays because the compression technology optimizes the transmission efficiency.

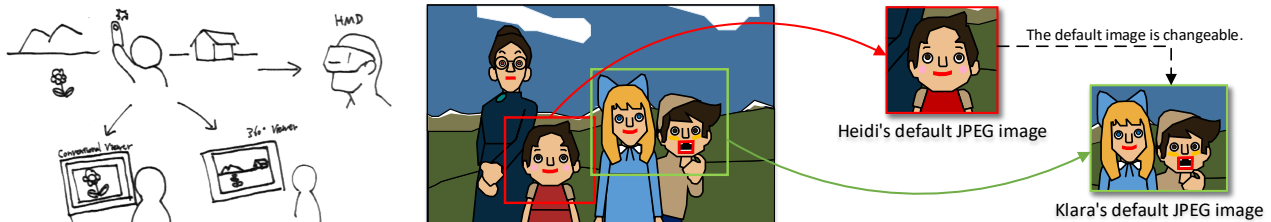


Figure 3.4 Navigating in Omnidirection Images

3.5. Adding additional media data to omnidirectional images

Mr. Urashima Tarō captures several pictures in a domed garden in the Ryugu castle with his 360 camera and also with a high resolution DSLR camera. He captures images of the same scene in the garden with different resolutions, and he also captures videos of a waterfall and a river which include audio. After collecting several images and videos, he embeds this image and video data into a 360 JPEG file, and he makes an omnidirectional image with multiple media data.

Tarō send his omnidirectional photograph data to Ms. Oto-hime. Oto-hime checks it with her 360 viewer (e.g. HMD, handheld device, etc.). She looks in all the directions of the omnidirectional image. She enlarges a particular area of his omnidirectional photo image, and she can see some zoom-upped area in a higher resolution where high resolution photo data exist. When she picks several image areas such as waterfall area of his photo, she watches a moving picture of that part and listen to the sound of it where the video and audio data exist.

Tarō stores several enhanced omnidirectional images in the cloud. It goes fast, because the files are optimized to reduce the amount of data transfer. Momotarō, Tarō's old friend, can review the photos without waiting for downloading of large amounts of data.

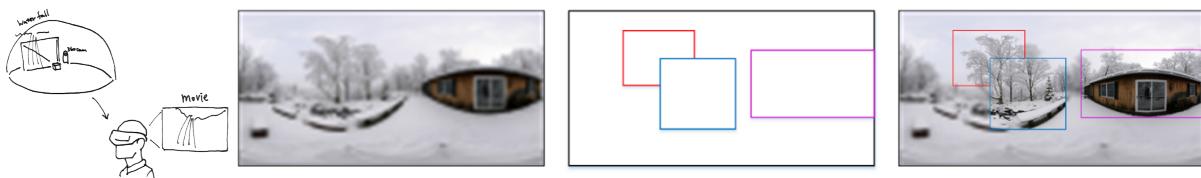


Figure 3.5 1.1. Adding additional media data to omnidirectional images

3.6. Time-Lapse Omnidirectional Animation

Ms. Rottenmeier captures a 360 picture of an aurora scene every five minutes. Then she makes an animated omnidirectional image of aurora. She also captures a 360 picture of a Japanese garden every day during the year, which allows her to compose an animated 365 days time-lapse image of the Japanese garden. After completing her collection of several omnidirectional scenes, she puts this data on her website.

Peter browses Ms. Rottenmeier's website by looking at the still photo images, and he decides to see her aurora's animated image data. After accessing the complete image data, he can play the animated aurora omnidirectional image with his HMD and 360 viewers on his PC and his smart phone. He enjoys Rottenmeier's omnidirectional animation of aurora.



Figure 3.6 Time-Lapsed Omnidirectional Animation

3.7. Computer-Savvy Tour Guide

Francesca earns a living by guiding walking tours through Padova; she's a native to Padova, and has been fascinated by its sights and history since she was a little girl. She offers an application that can be used by tourists for self-directed tours, and another application that can add labels and annotations to images captured by tourists after-the-fact. To create her catalogs, Francesca walked around with a 360 camera to the highlights she wants to capture, and then captures an image that includes the location data from her trekking GPS which is more reliable and accurate than the GPS in her phone.

When Francesca gets back home, she uploads her images onto her personal computer, and organizes them into catalogs for walks of different length of time, topics of interest (history, architecture, food & wine, restaurants, music, etc.); each image is tagged to at least one catalog, and often to several catalogs. Then Francesca opens each image and adds explanatory comments and image labels to call out the interesting facts about each place. The application then derives a signature for each of the tagged visual elements in the images, and using the GPS data, calculates positional data from which the tagged visual element was observed. Then Francesca adds links to her images, and links to additional online information and images for further details. Finally, the application extracts the positional data, image labels, image element signatures, and the Francesca's comments, and the additional links into a metadata set for each catalog that she offers, and copyright and authentication metadata is added. When a catalog is purchased, a customer-specific encryption is applied to the catalog; the key is provided separately.

3.8. Erudite Tourist

Gunther feels that wandering the city, while enjoyable, would be more enjoyable with the advice of guide, but Gunther dislikes large guided tours, and would rather spend his budget on a nice meal than getting personal tour guide. Lucky for Gunther, he noticed Francesca's self-guided tour, and he bought two of her catalogs: i) the two hour historical tour of central Padova, and ii) her latest guide to street food in Padova. Both catalogs contain very similar information, including visual signatures for the food served by the street vendors. Francesca's metadata catalogs are uploaded to his phone and his personal computer.

Gunther borrows a 360 degrees panorama camera from his friend François to use on his vacation in Italy. He straps it to his back and starts wandering Padova. The camera captures his walk by taking an image every 30 seconds, or when Gunther presses a button on a remote control in his hand in case something catches his interest. The camera also collects the time and position from the mobile phone Gunther is carrying, as well as the 6DoF and compass data from the camera sensor.

As Gunther walks, a mapping application shows him Francesca's tour path based on the GPS data from Francesca's file. He can stop anytime, take a side street, skip a sight, or stop for coffee, and then continue at his own pace. When Gunter finishes the walking tour, he uploads the images to his personal computer. Using the sensor metadata captured as we walked around, and combining with Francesca's metadata in her catalog, the appropriate metadata is carried over from Francesca's catalog into Gunther's images, though some of it must be recomputed due to positional

differences of the observations visual high lights. This merge/update of metadata become of Gunther’s images. It might also include Francesca’s copyright for her commentary. Gunther can decide whether or not to keep Francesca’s link to additional on-line information about the sights – probably will for now since he may want to learn more after he returns home from his vacation.

3.9. Multiple Resolutions Omnidirectional Images

Mr. Schmitz captures several 360 pictures of the scene in different resolution with his omnidirectional camera. The camera then fuses these images into an omnidirectional image file supporting multiple resolution omnidirectional data. Mr. Schmitz zooms in and out the omnidirectional image he has just taken, and he can see that his omnidirectional picture has functionality of zooming in any direction. He stores it on his web site.

Mr. Momotarō is having a look at Mr. Schmitz’s photos available online with an HMD viewer. Adolf’s web site offers low resolution photo image quality when browsing through his collection of omnidirectional photographs. Momotarō picks several photos and decides to access the full data of these photos. He can also display high resolution photos with his 360 degree viewing devices.

3.10. Layered 360 Images

Miss Orihime took many 360° pictures at almost the same time in many locations, and then she puts these all 360° pictures to as single format. Mr. Kintarō take a look to Orihime’s photographic image with his smartphone, and he find one house in her photo. He wanted to know what is behind the house, he use the transparent mode and select the house, and then the display application erased the selected house from the scene. After the house disappears, he can see a classic car behind it.



Figure 3.10 Layered 360 Images

3.11. Linked 360 Images

Mr. Kintarō checks the 360° image rooms on some real estate’s website for renting a house in the Shibuya area. He selects one picture and captures a look at the room with his HMD. He finds a door in the room. He clicks on the door, and his point of view moves to the next room which is a kitchen. He looks in several directions in the kitchen and he finds that there is a cabinet. After he clicks on the cabinet door, the cabinet door opens and he can check the cabinet inside.



Figure 3.11 Linked 360 Images

3.12. 360 Image Stabilization

Mr. Urashima Tarō creates a series of continuous scenes by continuously shooting 360° pictures of the Ryugu castle for several minutes in order to keep some memories. When Mr. Momotarō watches the pictures of the Ryugu castle taken by Tarō, he feels uncomfortable because the castle is shaking up and down. Mr. Tarō shot continuously without using a tripod, and so the Ryugu castle shook up and down. Fortunately, Tarō's photographs of the Ryugu castle has data of gravitational directions. Momotarō can correct the shake of the Ryugu castle easily by using that gravity data, so he won't experience any dizziness when watching Urashima Tarō's picture series.

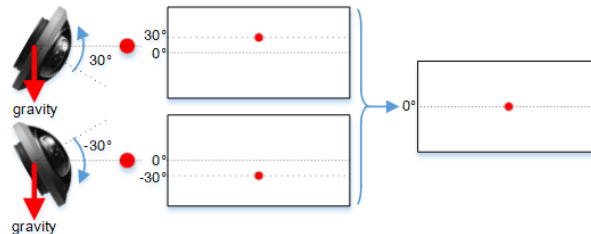


Figure 3.12 360 Image Stabilization

3.13. Efficient and Low Latency Viewport Transmission

Mr. Lee Min-Hyeong, a 360° image/video photographer, captures panoramic photos at the final race of women's 10 km speed skating relay at the Pyeongchang Winter Olympics with 8K (7680x4320) or higher resolution. After taking several picture sequences at the Winter Olympic Games, he stores this image data on his company's server, which delivers the data to their customers.

Ms. Jeong Yoo-Jin, one of the customers, takes a look at his sequence of speed skating relay pictures with her HMD. The display system has been created assuming a viewing angle of 90 degrees. The viewport has 2K resolution (1920x1080) and it is transmitted in high quality from the server side to HMD, while the data around the viewport is blurred. It becomes possible then to receive clear image at high speed. This technology allows Mr. Jeong Yoo-jin to view comfortably the Olympic Winter Games photos and videos.

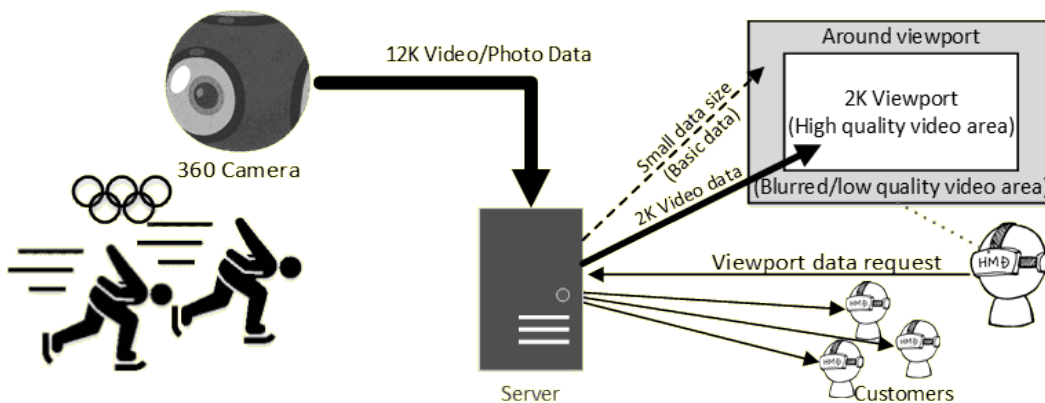


Figure 3.13 Efficient and Low Latency Viewport Transmission

3.14. Guided Viewport Experience

Mr. Kitarō takes several spherical photographs, and he decides to add information to the photograph about displaying it in his favorite way. With this information other people can see his photograph in his recommended viewing way. He also puts three thumbnails of his most favorite scenes to the captured spherical image for instant reviewing with web browser, so everyone can immediately access his favorite scenes.



Figure 3.14 Guided Viewport Experience

3.15. Virtual Museum Tour with Explanatory Materials

Miss Neko-Musume goes to the web site of the National Museum to see modern art from her convenient location. Using her HMD, she goes through and see several pieces of art in the rooms of the museum via real/virtual corridor which connects the rooms. One picture catches her eyes. She wants to know what this painting art is, and she decides to display a detailed commentary of the picture. Then an explanatory note appears on the wall next to the picture, and she is able to know more about the painting art.

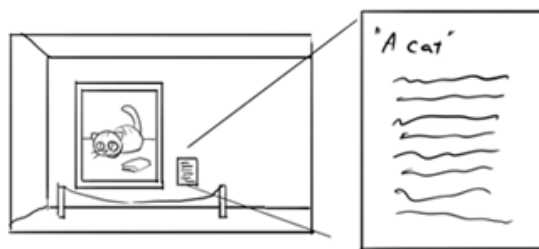


Figure 3.15 Virtual Museum Tour with Explanatory Materials

3.16. Camera Position and Orientation

Mr. Kitarō captures several photographs using his 360° camera which stores its shooting position, direction and time. With this data he can make a relative map easily. And he is also able to see which photograph is taken at which position and when.

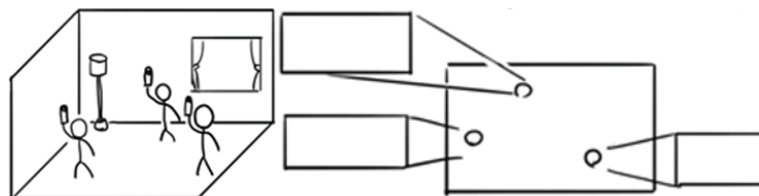


Figure 3.16 Camera Position and Orientation

3.17. Selective Filtering Indication

Harry buys a smartphone with a 360 camera. In order to draw visual attention to a portion of the 360 image, Harry can define a masking region of interest ROI when editing his 360 photos such that where the mask is defined to overlay the captured image, a visual blurring effect is applied. This allows Harry to send the image + mask to his girlfriend Sarah to put the attention to the interesting part. If, however, Sarah wants to see what was in the background without the blurring effect, she can disable the effect and see the original captured image.

3.18. Stereoscopic 360 Image

Frank captures an image with a stereo camera system during his vacation. He put the image on facebook, so that his friends can check his visited location. His friend Harry wants now to have a more immersive impression of the location, so he puts on his HMD to watch the image in stereo.

3.19. Multiple Type Sensor Images

Miss Orihime has a smart phone with dual type image sensors which can capture coloured image and infrared image. These image sensors are arranged in parallel so that the same object can be taken at the same time. Her smartphones can make wonderful photos from these images, however she saved these sensed image as one file format, and she apply other algorithms and sent it to the web service.

3.20. In-image Face Tagging for Photos

The photographer Jean Stringer takes a 360 degree photo for the photo department of the news agency AllNewsNow at a meeting of the heads of the 28 EU countries in Brussels. The 26 persons present at the time of the shot are seated around a table of elliptical shape and the 360 degree camera is placed in the middle of it. (Note: the illustration image does not meet this position, it is placed at the edge of the table.)

The photo editor of AllNewsNow has to add the names of the heads-of-state as metadata. As the typical viewer software of 360 degree photos opens only a limited viewport to the spherical view the photo editor wants to associate the name of a person with her or his face individually. First he draws a virtual rectangular numbered frame around the face of each head-of-state looking towards the camera in the 360 degree photo.



Figure 3.20.1 Captured Photo with Face Position Indications

Then he adds the name to the Person Shown metadata field of each numbered frame.

AllNewsNow Photo/Video Metadata Editor's Desk

Frame	Metadata of that frame
#1	Person shown: Ms Aname
#2	Person shown: Mr Cname
#3	Person shown: Mr Hname
#4	Person shown: Mr Kname
#5	Person shown: Mr Lname
#6	Person shown: Mr Mname
#6	Person shown: Mr Tname
#8	Person shown: Mr Uname
#9	Person shown: Mr Zname
#10	

Figure 3.20.2 Association of Person's Name to Frame Number

The information of each frame is transformed into an individual set of XMP metadata and all sets are embedded into the JPEG photo file.

A client of AllNewsNow launches a 360-degree-photo-viewer software to have a look at this JPEG photo. While moving the viewport across the sphere she can click on the M key or touch the screen of the tablet or smartphone and the viewer software shows the names next to the faces which are currently shown:

View in one direction ...



Figure 3.20.3 FOV in One Direction

... view in another direction:



Figure 3.20.4 FOV in A Second Direction

... view in another direction:



Figure 3.20.5 FOV in A Third Direction

This definitely helps to distinguish which faces out of the many heads-of-state around the table are currently shown in the viewport and this improves the acceptance by consumers, and this improves the business value of 360-degree photos with metadata for different regions inside it.

AllNewsNow highly appreciates a very similar way of editing metadata for frames inside a 360-degree video.

IPTC considers it as key technical requirement for metadata associated with virtual frames to have a standardized reference point or reference coordinates system for a 360-degree spherical view. This is the origin for the definition of other points in this sphere relative to it. These points can be used to define such virtual rectangles.

Requirements

- To specify the relative location of tagged image elements a standardized origin is needed.