“JPEG XT” A New Family of JPEG Backward-Compatible Standards

Thomas Richter (richter@tik.uni-stuttgart.de),
Alessandro Artusi (artusialessandro4@gmail.com),
Touradj Ebrahimi (Touradj.Ebrahimi@epfl.ch)

(This is a preprint of the article that appeared in the July/September 2016 issue of the IEEE Multimedia Magazine)

"JPEG" is today a synonym for "Digital Photography", and "the format in which digital cameras store pictures". But it sometimes comes with the connotation of "quality loss" and "precision loss", even more so as camera vendors equip their products with the ability to encode pictures in proprietary formats, typically called "RAW".

"JPEG" is not only the name of a still image format. It is also the name of a joined standardization committee under auspices of the International Organization for Standards (ISO) and the International Telecommunication Union (ITU-T), formally known as ISO/IEC JTC1/SC 29 WG 1. In fact, JPEG stands for Joint Photographic Experts Group and collects, assesses and validates technologies and recommends imaging international standards. As such, during more than two decades in existence, the JPEG committee has developed many more standards than just the widely used legacy JPEG format (ISO/IEC 10918-1 or Recommendation ITU-T81). Among these standards one can mention JBIG and JBIG2 - two bi-level compression formats used in transmission and storage of black-and-white prints the average photo-copier speaks - and JPEG 2000, the format that cinemas use nowadays to digitally store and transmit movies shown in theaters.

Even though better in performance and offering many more features, JPEG 2000 has not been adopted in consumer digital camera market and did not replace the widely used legacy JPEG format. It is often claimed that JPEG 2000 has a complexity issue, though when compared to today's video coding standards, e.g. HEVC (also known as ITU-T Recommendation H.265), it is rather lightweight!

The problem actually lies elsewhere, namely in interoperability. JPEG 2000, or even another more recent and less complex JPEG format known as JPEG XR, are not backward compatible with the widely used JPEG format. Concretely speaking, the JPEG decoders in your PC, your TV or your electronic picture frame won’t be able to decode such more recent formats. Even though the performance of JPEG 2000 or JPEG XR is superior, it does not seem convincing enough for investment in new equipment, both in hardware (camera, picture frame, TV, etc.) and software (image viewers, image editing apps, etc.).

Surprisingly, words such as "lossy" and "low precision" typically attributed to JPEG are not even fully correct as the legacy JPEG format also includes modes for lossless encoding and higher precision encoding. Unfortunately, even though part of the same standard, these modes are not compatible with the widely known lossy 8-bit mode of legacy JPEG format that became so overwhelmingly popular.

To address the above needs, the JPEG committee has recently produced a new standard, JPEG XT, or formally, ISO/IEC 18477. It is both backward-compatible to the legacy JPEG, and offers the ability to encode images of higher precision, higher dynamic range, in lossy or lossless modes. Any legacy JPEG decoder will be able to decode a JPEG XT file, and in that sense, JPEG XT is even more compatible than JPEG ever was, as it leaves all the less popular modes out. Clearly, if your decoder only understands JPEG and not JPEG XT, it will only get an 8-bit lossy image as it would expect. Lossless decoding, or full sample precision, would still require a full JPEG XT decoder.

JPEG XT does not stop there. JPEG XT offers several other features that go beyond those in legacy JPEG. Some of them are under standardization, such as the inclusion of opacity information (alpha channels), while others are in the pipeline of the JPEG committee. These features include privacy and security protection of image content or image regions, animated JPEG, and efficient coding of omnidirectional (also known as 360-degrees) images.

In this article, we shed some light on the recent developments of the JPEG committee by starting with some basics on the current JPEG standard and then proceed to JPEG XT, its current status and its future plans.

JPEG is not quite the JPEG you know, and “RAW” isn't as raw as you think.

JPEG compression is based on a single mathematical trick: the original image is first separated into 8x8 pixel blocks, and each block is then transformed by the so-called “Discrete Cosine Transform - DCT”. DCT has the interesting property that the transformed data consists of a single large coefficient representing the average luminance of the block, and many remaining smaller values representing the structure in that 8x8 pixel block. For most natural images, the structure information is simple and can be approximated with just a few DCT coefficients without causing visible distortions.

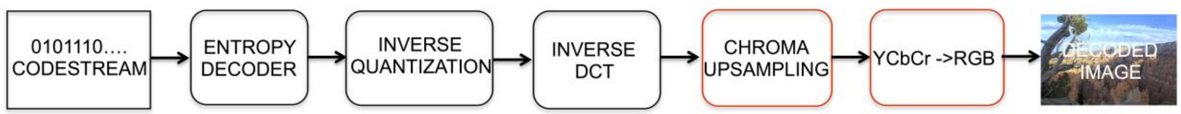


Figure 1: The JPEG decoding algorithm in a nutshell

The legacy JPEG format is often associated with a compressed image limited to a precision of 8 bits per component, i.e., 256 different values per color channel. JPEG images also use a conversion from an RGB space (Recommendation ITU-R 601) to YCbCr to decorrelate color components, and use subsampling in the chroma signals - much like analog TV did - to further compress the image.

The surprising fact is: these commonly known attributes of JPEG are not even part of the JPEG standard. They are part of JFIF, a proprietary solution that has found wide adoption and that specifies the color decorrelation and component subsampling mentioned above. The JPEG standard itself only describes how to encode numerical sample values on multiple rectangular sampling grids, possibly related to subsampling.

Neither is JPEG confined to 8-bit coding. The original JPEG standard also defines a high-precision 12-bit mode, and a lossless mode, and a few additional modes that have not found wide adoption and popularity. One may speculate why these modes never became popular, but the fact is that they are not compatible to the widely used 8-bit mode, i.e., a decoder only implementing the latter is not able to decode images in other modes, not even with reduced precision. This is however different for JPEG XT.

JFIF was only standardized lately as ISO/IEC 10918-5 (JPEG Part 5), and JPEG XT in its Part 1 also adopts the same procedures. As opposed to the original JPEG standard, JPEG XT Part 1 does not adopt all the little-known modes that did not gain in popularity. In that sense, JPEG XT Part 1 is much closer the JPEG format we all know than the actual original JPEG standard [8].

Digital camera vendors currently address the needs for higher bit depth and higher precision by a format called "RAW". Unfortunately, "RAW" is not really a format, and not quite as raw as one might wish. “RAW” does not refer to a single format because every vendor defines its own way of coding images captured by their cameras, so the “RAW” format from vendor A is not compatible with the “RAW” format from vendor B. The “RAW” format may even change between camera models of the same vendor year by year. In worst case, you might loose all your pictures captured in “RAW” if the camera vendor decides to not to upgrade their “RAW” decoder for your old camera model to the latest version of your operating system.

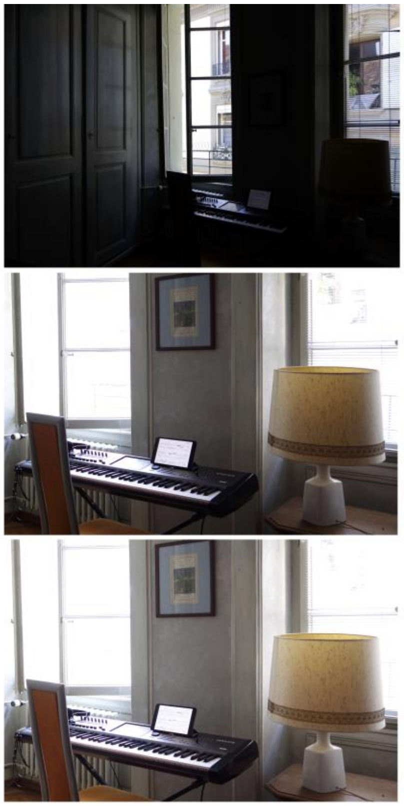
Unlike claimed, "RAW" is neither an unprocessed data. It seems tempting to believe that the sample values of the decoded raw image are the unmodified sensor reads from the camera sensor. This is, however, typically not the case. Depending on the vendor, the camera may perform several preprocessing steps on the sensor data. It is amplified according to the ISO level setting of the camera, then digitized and probably some dead reads or dead cells are removed. Some cameras already include flare removal at this stage. What you get as raw depends very much on the design decisions of the vendor.

Higher Bit Depth and High Dynamic Range

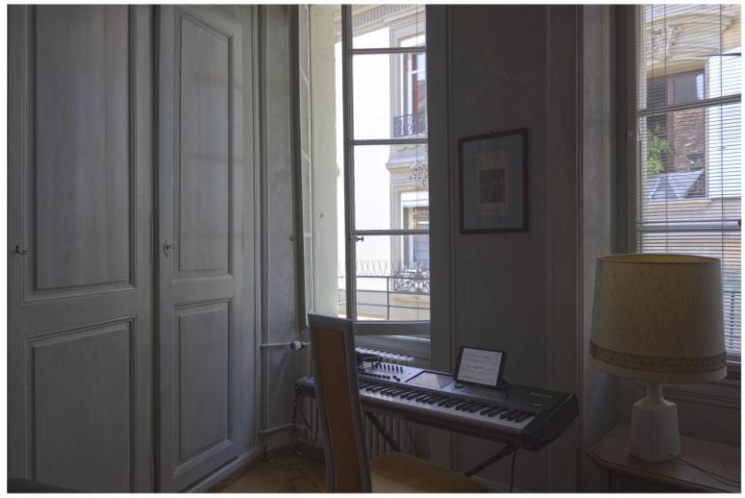
The first backward compatible enhancement brought to legacy JPEG by JPEG XT – "XT is short for eXTension" – is to enable coding of images with a bit depths higher than 8 bits per component. Unlike an image coded in the little known 12-bit mode of JPEG, a legacy JPEG decoder only implementing the popular 8-bit mode would still be able to understand a JPEG XT codestream, though the output would then only have a precision of 8 bits per sample. JPEG XT does this by first encoding an 8-bit version of the high-precision input, also called base layer, and hiding a second codestream known as enhancement layer, within this legacy codestream that enlarges its precision to a fuller range; up to 16 bits per component or 48 bits in total as currently specified. Additional metadata, also embedded in the legacy codestream, tell a JPEG XT decoder how to combine the base layer and the enhancement layer to form one single image of a higher precision.
Embedding mechanism used in JPEG XT is possible thanks to a legacy JPEG structure called "application marker". The original intent of these markers was to allow applications to add vendor-specific information to the codestream without breaking compatibility. Decoders unaware of a specific application marker would ignore it and read over it. In the case of JPEG XT, these markers are used to hide JPEG-XT related information from legacy JPEG decoders.

More precisely, the procedure mentioned above works as follows: In a first step, the 8-bit image in the legacy codestream is mapped into the full output precision by a freely specifiable, but global function, that assigns to each 8-bit value an approximate high-precision value. The second image invisible to legacy JPEG compliant implementations, decodes to an error signal that is added to the approximation to form the full precision output. How exactly this function looks, and how it has to be picked is not specified in the standard. In the simplest possible case, it might be an upshift, i.e., a multiplication by a power of two. The least significant bits come in this simple example from the second "residual" image.
In practice, higher precision images are often encoded without gamma correction, i.e., the sample values are proportional to physical intensity rather than proportional to a power function of the intensity, whereas 8-bit images typically make use of such a gamma correction. In such a case, this additional non-linearity has to be corrected when mapping from the 8-bit input to the high-precision output.

If the range of sample values in the image is so large that it does not allow description by 16-bit integers, we're entering the domain of high dynamic range (HDR). Here, images are typically obtained by taking multiple shots of the same scene while varying the exposure time (see Figure 2). Software can then compute from these inputs one consistent image where each sample value is proportional to the physical radiance in the scene [1]. For natural outdoor scenes, the radiances can vary over four to six orders of magnitudes, too large to capture by one single typical sensor or to allow representation by a 16-bit integer.



(a) Three LDR images taken at three different exposure times



(b) HDR Image obtained fusing the three LDR images in (a) and tone-mapped for display.

Figure 2: Multi-exposure approach to acquire HDR image. Different shot of the same image taken at different exposure time are merged to reconstruct the full luminance dynamic range of the real-world scene.

Multiple picture formats have been developed for this particular use-case. RGBE [2] stores, for example, for each red, green, blue triplets, i.e., for each sample, one additional common exponent by which the three values are scaled. OpenEXR [3] is an HDR format for movie post-production, developed by Industrial Light & Magic. It stores for each pixel the radiance of red, green and blue in a 16-bit floating-point representation, and hence uses separate exponents. The well-known TIFF format also allows a similar encoding, though offers an even larger precision of up to 32-bits floating point per channel.

What is common to all three formats is that they are uncompressed, or use a very simple compression algorithm with limited efficiency. In particular, HDR images in such formats become particularly large. JPEG XT uses here the same paradigm as for encoding of high-precision images. It includes an 8-bit version of the HDR image as base layer, which a legacy JPEG decoder can recognize, and an extension layer that enhances the dynamic range to the full dynamic range of the original image. A JPEG XT codestream again includes additional metadata about how to assemble the base and enhancement layers together. The 8-bit image visible to legacy decoders is here typically formed by "tone mapping" the original HDR image to a low dynamic range version.

Tone mapping is a compression of the dynamic range that is, however, typically adapted to the image characteristics to avoid loosing contrast and details, e.g., edge information, in the original image. There is no single correct recipe on how to perform such a tone mapping; it is often an artistic choice the author of an image makes, and for that reason, many tone-mapping algorithms have been suggested in the past. JPEG XT does not enforce any particular choice here; the input to an encoder consists of an image pair - an HDR image and its tone-mapped 8-bit version.

JPEG XT allows great flexibility in the specification of the algorithm that assembles the HDR image from the tone-mapped 8-bit image - the base layer - and the extension layer. To manage complexity in decoders, it defines four Profiles - subsets of the full feature set - each of which uses a relatively simple algorithm to merge base and extension layers (see Figure 3).

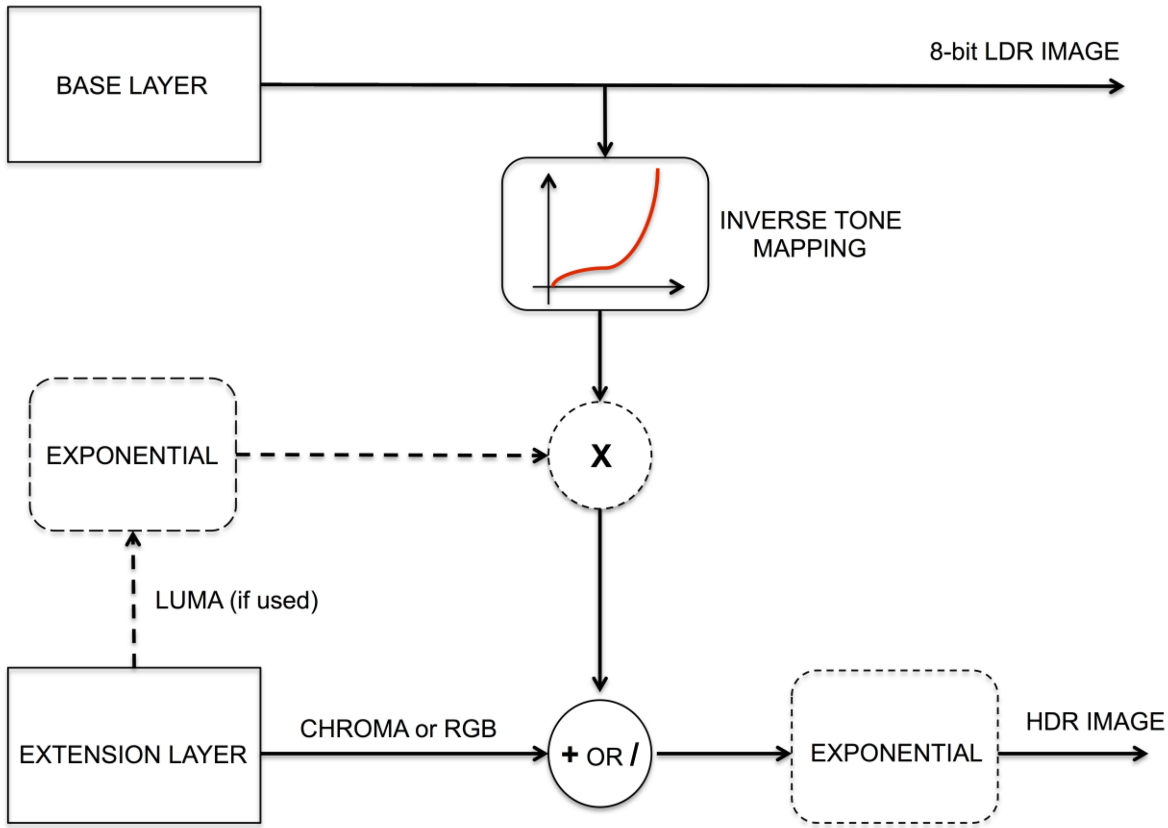


Figure 3: JPEG XT decoder structure provides a unique architecture for different profiles. The above diagram shows a common architecture for profiles A, B and C.

In Profile A, the HDR image is described as the pixel-wise product of an inversely gamma-corrected 8-bit tone-mapped image multiplied by a common scale factor, not quite unlike the RGBE format. The scale factor is encoded in logarithmic space, i.e., it is also an exponent as in the RGBE representation. In addition to the RGBE format, Profile A also includes an additive chroma residual that allows correction of color defects that might arise due to simple scaling [4, 5].

In Profile B, the HDR image is described as the pixel-wise quotient of the gamma-corrected tone-mapped image and a divisor image delivered by the extension layer. The output is additionally scaled by one common factor, the "exposure value". This profile is motivated by the following observation: assume that the base layer consists of a gamma corrected version of the original HDR image that was clamped to the luminance range available to legacy JPEG. As long as the pixel values in the HDR image are within this range, the denominator represented by the extension layer may remain one. If the luminance in the base layer was clamped to the maximum possible value and the HDR luminance is not representable by means of an 8-bit format, the denominator must fall below 1.0 to amplify the output of the JPEG XT decoder to the sample value in the original image. Hence, for such a simple tone-mapping, the extension layer contains all the overcast image regions not carried by the base image, and the base layer contains all regions that are visible under regular exposure. The exposure factor then controls where the cut between regular and extended dynamic range lies [4, 5, 6].

Profile C works almost like Profile A, i.e., it decomposes the image into an 8-bit image and scale factors, except that here each component has its individual scale factor - unlike in Profile A which only allows for a single common scale factor. Profile C also performs the multiplication as an addition in logarithmic space. An additional trick in this profile is that the logarithms and exponential functions to map between linear and logarithmic space are approximated by piecewise-linear functions that have exact inverses, enabling lossless coding [4, 5,6].

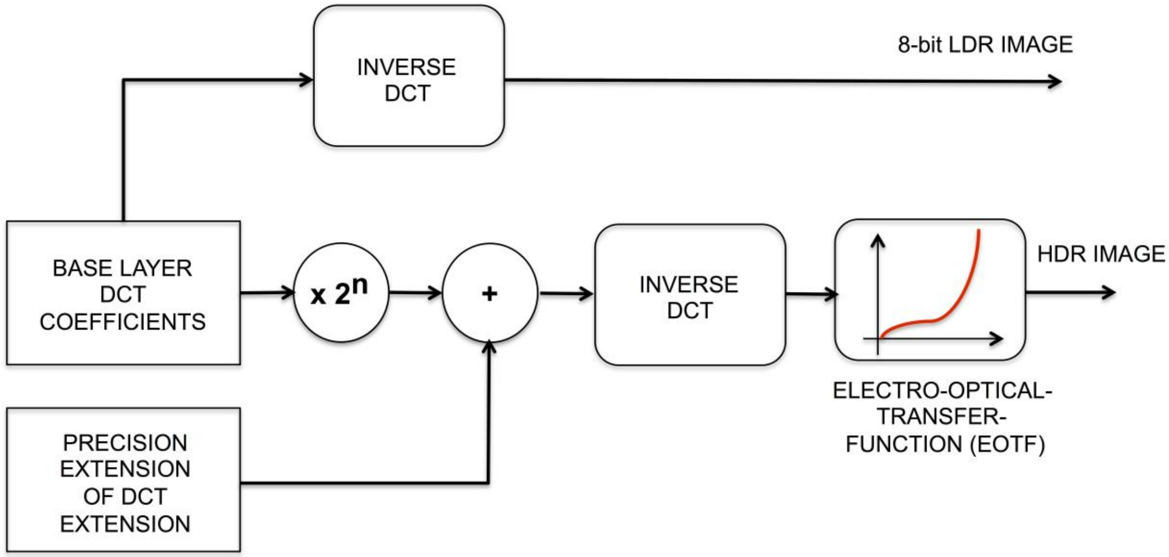


Figure 4: Profile D overview.

Profile D is somewhat different from all previous profiles in the sense that it does not include two separate images. While Profiles A to C first reconstruct base and extension layers as independent JPEG streams and merge them pixel by pixel in the spatial domain, Profile D extends the resolution of the image precision by extending the precision of the DCT coefficients beyond the resolution allowed by legacy JPEG [8]. The reconstructed image data then undergoes an "electro-optical transfer function" (EOTF) signaled in the codestream, an idea related to the gamma transformation used in standard dynamic range.

Due to the non-linearity of the human eye, luminance differences at high background intensities are less visible than luminance differences in dark environments, resulting in an approximate logarithmic sensitivity curve. This sensitivity curve can be exploited by compressing brighter image regions more than darker image regions, allowing a faithful HDR image reproduction from 10 or 12 bit samples. A very similar idea, namely that of an EOTF, is also part of the MPEG HEVC "range extensions" for compression of high dynamic range video.

Similar to Profiles A to C, Profile D offers backward compatibility, but only to a limited degree. Profile D files are readable by standard JPEG decoders, but due to the absence of the EOTF in legacy implementations, colors may appear dull and the overall image may become unappealing, though visible. The extension data allowing 10 or 12 bit samples is again embedded as an extension layer [6].

JPEG XT also allows mixing of profiles, so some elements from Profile C can be mixed into a Profile B configuration, or the Profile D extension of the precision of the DCT samples can improve the extension layer in Profile C [7]. The resulting codestreams are then called "full profile" codestreams because their configuration exceed the constraints implied by a single profile, though such streams are fully covered by the JPEG XT standard [6].

Lossless Coding and Lifting

Even thought the legacy JPEG standard also defines a lossless mode, it has not found wide adoption - except in some medical applications - and JPEG XT Part 1 does not include this legacy coding mode because it is not compatible to the popular DCT mode. Furthermore, due to the simplicity of the lossless mode - Huffman coding of the prediction error of horizontal, vertical or diagonal neighbours - its coding efficiency is neither very competitive. The JPEG committee in fact has produced two standards with JPEG-LS and JPEG 2000 based on two smarter algorithms that also include lossless modes; JPEG-LS is also based purely on coding of prediction residuals, but is much more sophisticated than the lossless mode of JPEG. Due to its simplicity, it has found applications in the transmission of images from the Mars Rover and in satellite imaging applications.

None of these codecs are, however, compatible to JPEG, and existing JPEG decoders cannot even reconstruct a lossy version of the original image from a JPEG-LS or JPEG 2000 codestream, quite unlike JPEG XT. The latter offers approaches to enable lossless coding: if 8- to 12-bit precision is sufficient, a lossless integer-to-integer DCT replaces that of legacy JPEG which is defined only loosely. Similar to conventional DCT, the lossless DCT is split into a sequence of "elementary rotations", i.e. linear transformations that operate on pairs of coefficients by rotating them in the $\mathbf{R}^2$ space by a step-dependent angle. The lossless DCT replaces each rotation by an equivalent sequence of three shearing steps, each of which is invertible; this trick, also known as "lifting", is also the building stone of the wavelet transformation within JPEG 2000. While a rotation is only invertible up to an implementation specific precision loss, each shearing depends only on integer operations and is hence exactly invertible.

The second approach to lossless coding in JPEG XT also requires to fully specify the DCT process. Unlike the first approach using a lossless DCT, it is based on a lower-complexity fixed point based approximation of the DCT. This DCT is, by itself, not completely lossless [6]. Similar to the lossy modes, a second codestream - hidden within application markers – then includes the residual error to compensate for the predictable coding error of the DCT in the base layer. Because the DCT is fully specified, an encoder can predict the errors cased by DCT in the decoder, and is able to compensate them to allow lossless reconstruction.

A legacy decoder will only recognize the base layer and will reconstruct it by its lossy DCT algorithm. Lossless reconstruction requires a JPEG XT conforming decoder. For that, of course, the error residuals need to be decoded without loss, something native JPEG cannot ensure. Two possibilities exist: Either to simply bypass the DCT in the extension layer and encode directly in the spatial domain without further transformation, or to transform the error residuals with the lossless integer-to-integer DCT of the first mode.

Coding of Opacity Information

One reason behind popularity of PNG image format in web graphics is that it supports the inclusion of opacity information, i.e., the definition of an additional mask that defines how transparent a given pixel in the image is. A common trick one can play with transparency is to use it to define graphics objects with arbitrarily shaped boundaries that embed smoothly into a background - or text that floats around.

Unfortunately, the legacy JPEG standard never defined how to include opacity information, so web designers have to sidestep to PNG if this feature is required. Unfortunately, PNG does not exhibit a good compression efficiency when compared to JPEG, so bandwidth requirements go up as soon as opacity information becomes necessary.

JPEG XT allows the possibility to include a transparency layer by the same means it includes extension information for higher bit depth: a second codestream, decoding to the opacity information, is hidden within the application markers of the actual foreground image, along with the meta-information that describes how to decode it. Hence, the image in total consists of two JPEG codestreams: The foreground image, and hidden within it, a one-component JPEG image decoding to the alpha channel [6] (see Figure 5).

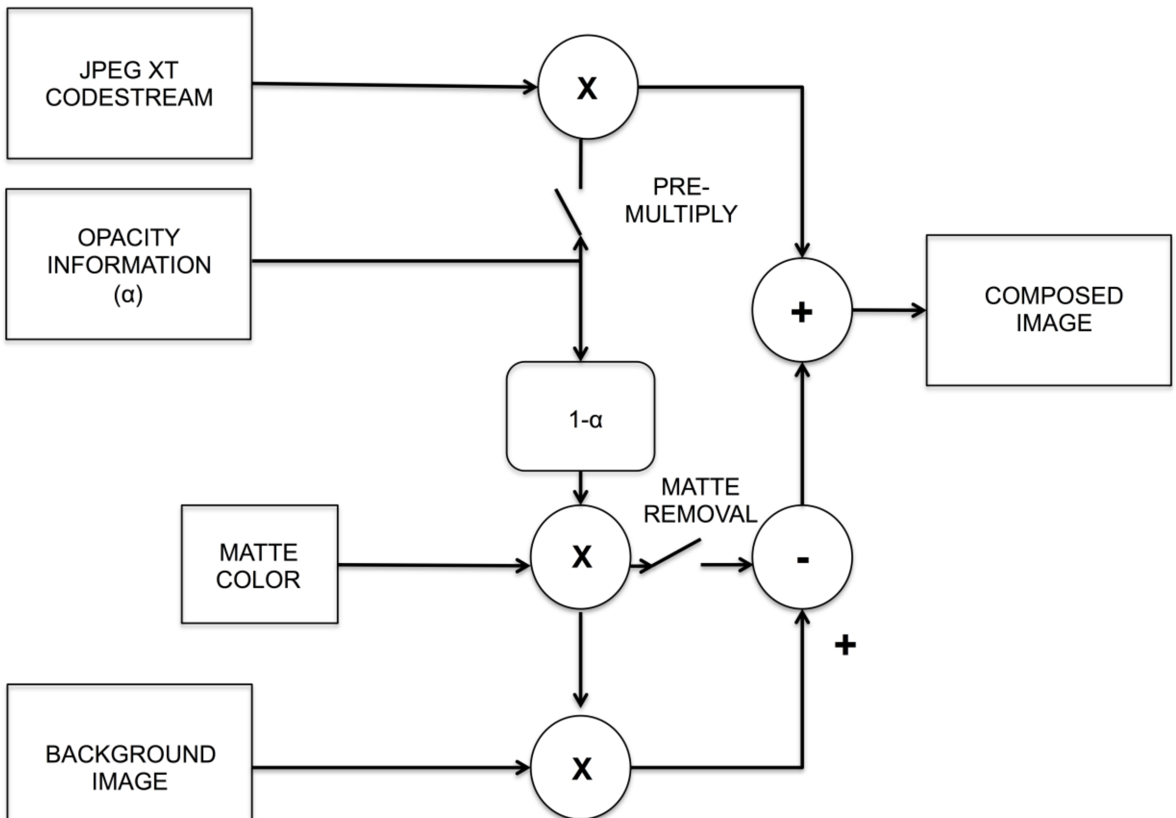


Figure 5: Alpha channel JPEG XT decoder.

While 256 levels of opacity, as granted by a legacy JPEG 8-bit description, is sufficient for many applications, the full feature set of JPEG XT is available to encode the opacity information. Hence, even floating point based opacity samples with a precision of up to 16-bits can be carried in the extended format.

As for all JPEG XT specifications, an image including an opacity information is decodable by any JPEG decoder, though only decoders conforming to JPEG XT will pick up the transparency layer.

Privacy protection and security in images and image regions

Millions of images are shared today on social networks, most of them in the legacy JPEG format. While it is typically possible - within limits - to restrict access to the picture to a selected group of individuals, a given image can only be shared in total, or not at all. Whether your friends on the picture want to be visible, or to whom they want to be visible, is not in their hands. Once a picture is taken and has been uploaded to an image repository, access control to its content has been handed over to the repository, and you as the photographer have to trust the management of this repository that it will handle your images as promised.

JPEG privacy and security is a standardization initiative that puts the control of image distribution back into the hands of the user. While details on the technology are still being worked on, the JPEG committee already identified a list of requirements a future JPEG privacy standard needs to address: JPEG privacy and security will not only allow protection of an image as a whole, it will also allow to constraint access to image regions. A very typical application is to pixelate only the faces while keeping the rest of the image intact. The resulting privacy-protected image could be distributed without further restrictions, while access to the full image, i.e., the image with recognizable faces - would be limited to a close group of friends.

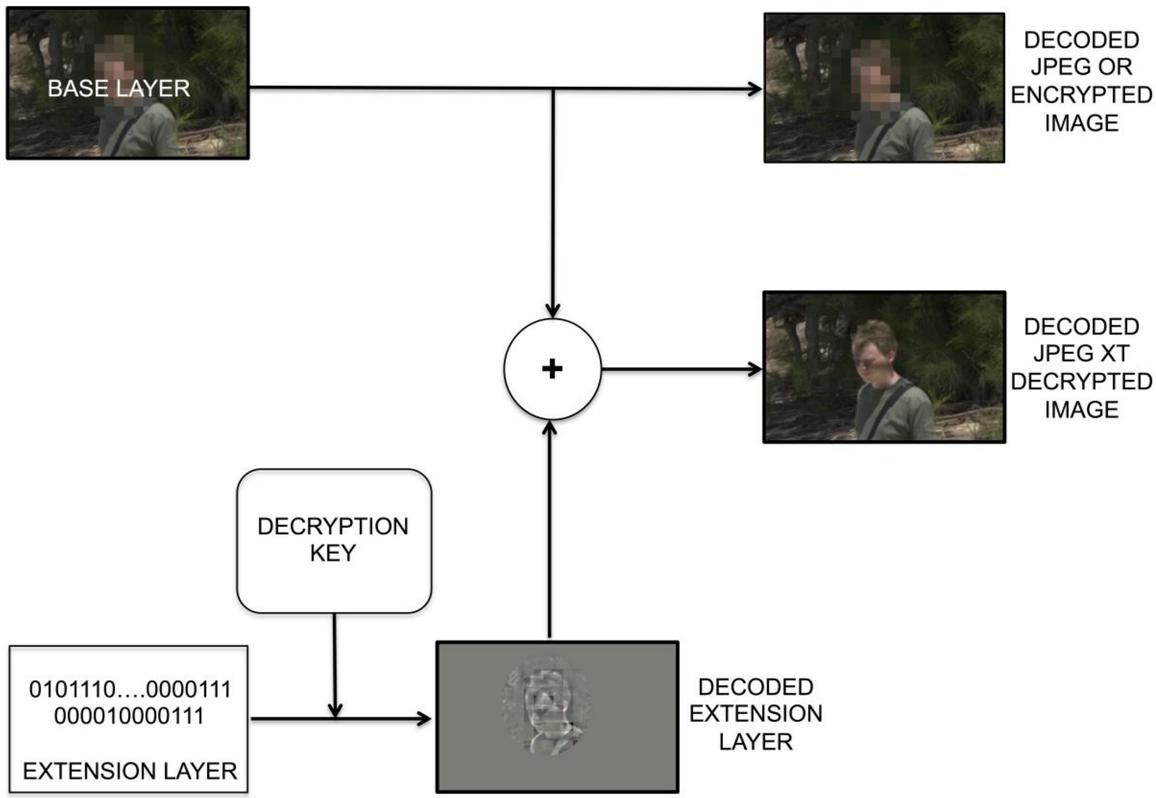


Figure 6: JPEG XT privacy extension (Image courtesy of Olivier Pfeiffer).

The envisioned technology for JPEG privacy and security consists of multiple parts: First, the original image needs to be segmented into public regions that can be shared without restriction, and image regions that require privacy protection or other security mechanisms. Such a segmentation can be done either manually, or automatically, e.g. by face detection algorithms.

Second, the identified image regions need to be pixelated or protected by some means. At the same time, an encryption process may need to generate sufficient data to identify the modified regions, plus the data allowing descrambling protected regions. A simple algorithm would, for example, compute the differential image between the original and the scrambled version. This differential image would be zero in all regions that are identical in the original and scrambled image, and would contain the pixel-wise difference between the two versions otherwise.

Third, the data required for decoding the original image, i.e., the differential image, is encrypted by a standard public/private key algorithm. It is important to select here a proven and robust algorithm, and JPEG will certainly not define its own encryption process.

Last but not least, the encrypted data needs to be embedded into the codestream of the scrambled image. JPEG XT already defines means on how to hide such additional image and metadata from legacy decoders, and how to structure them; JPEG privacy and security will very likely build on the same technology used in JPEG XT to store the extension data for HDR photography.

Interestingly, JPEG privacy and security is not the first standard that identifies privacy and security protection of digital images as an important feature: In 2007, the JPEG committee developed JPSec [7], a standard that allows protection of JPEG 2000 codestreams. JPSec was not accepted by the market; whenever protection of JPEG 2000 images became relevant, for example in digital cinema or medical applications, the protection tools and frameworks of the corresponding application layer were deployed instead of JPsec, and there was no need to integrate privacy and security protection closely into the image format. Instead, the image codestream was protected as a whole. This simple solution is, however, no longer applicable if only parts of an image should be protected.

Potential Extension of Features of JPEG XT Standard

The layered structure of JPEG XT provides a powerful tool to extend the standard capabilities into various novel applications. Some of them have been presented in the previous sections and are under development by the JPEG committee. Others are under consideration or part of potential future plan of the JPEG committee: omnidirectional photography, structured image editing and animated JPEG are among them.

Omnidirectional Photography

Omnidirectional photography also referred to as 360-degrees panoramic imaging is a digital image where all the scene information around the camera is captured. With proper viewing devices, such as the emerging VR technology, omnidirectional images allow the total immersion of an observer into a scene.

Most 360-degrees panoramic images are captured through a process known as stitching which consists in a series of images assembled together, where each image is from a different view. A potential extension of the JPEG XT format can be used to store this type of images efficiently. As in all of JPEG XT, backward compatibility is a major goal. One possibility to realize this form of compatibility is to select an interesting window within the omnidirectional content, either automatically or manually, as the main view decodable by legacy JPEG decoders, i.e., this particular view defines the base layer; while the entire omnidirectional content will be available for a full decoder by means of an extension layer.
An alternative design might represent the entire view as a geometrically distorted JPEG image, and the JPEG XT codestream would only include the necessary information describing the geometry of the scene, and hence to correct geometrical distortions dynamically depending on the view angle of the observer.

(a) Full Omnidirectional Photo; with region of interest (red)

(b) Region of Interest as backward-compatible image

Figure 7: Panoramic images within JPEG XT.

Structured Image Editing

The editing process of a digital image may be tedious and time consuming; it may hence be desirable to provide a user-friendly image format that contains all the previously edited versions of the original input image or images, plus all the editing information that were necessary to generate the output. Such a format may not be only useful for a digital artist; together with appropriate digital signatures that may become part of the JPEG privacy and security initiative, it may also aid to identify the authenticity of images and enable users to follow the process flow of an image from the capture to the pre-press.

The layered structured of JPEG XT is suitable for recording the editing history of a digital image. The latest version of the edited image, carrying the final artistic touch can be stored in the base layer and will therefore be available to all the JPEG decoders, regardless whether they support or not the JPEG XT standard. The extension layers will carry all the earlier versions of the edited input image enabling the ability to travel back in time. This information will be available only to the JPEG decoders that support JPEG XT standard [5].

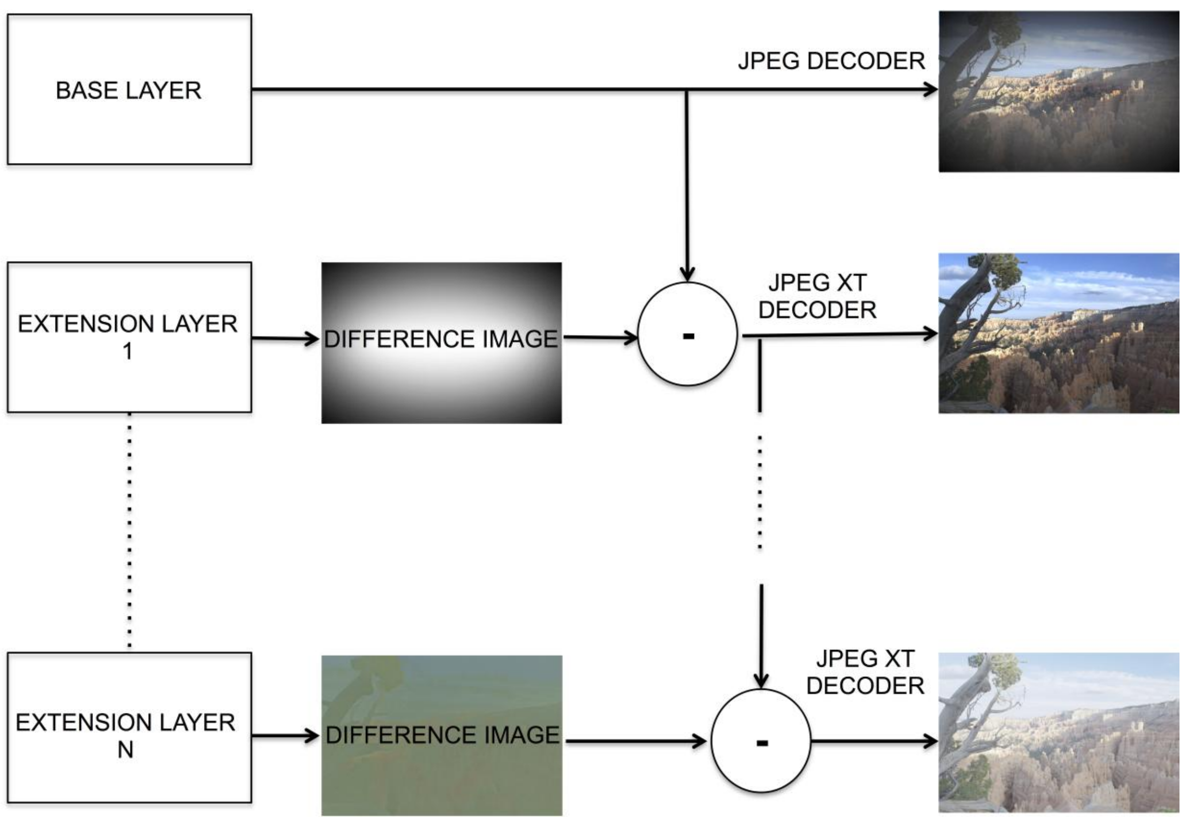


Figure 8: Storing the image editing history in JPEG XT.

Animated JPEG: JPEG as a flip-book

Long before PNG became popular, GIF was the dominant format for small icons and buttons on web-pages. GIF also supported one particular neat feature: Similar to a flip-book, a single GIF file could include more than one image, forming an image sequence over time. Each image includes a duration for how it is to be shown, and a flag defines whether to loop back to the first image at the end of the sequence.

Animated GIF was never intended to replace full-fledged video formats like MPEG; instead, it was used to pimp up webpages by small animations like rotating buttons to catch the attention of the reader. The PNG format later on took over, and also added the possibility for animated graphics.

In the JPEG world, animation was standardized - if at all – relatively late. JPEG 2000 includes the possibility to combine several images in an animated sequence, though this extension is specified in Part 2 of the format (ISO/IEC 15444-2) which is less popular, so it did not gain much attention.

While animated JPEG does not yet exist as a standard, Motion JPEG is a particularly popular format; in specific, simple web cams use it to stream their images over USB to the host system. Technically, it is just a sequence of individual video frames, each of which is an encoded JPEG image, though it has never been standardized and lacks any framing of itself - or rather, it is typically used as a codestream in a file format in an application-dependent way.

The use case of Motion JPEG is orthogonal to that of animated GIF. Motion JPEG is a very basic video format and as such replaces the more complex MPEG formats for cheap and simple applications where images are small enough to make bandwidth considerations irrelevant. Animated GIF is not intended to play longer video sequences, but short animations as in a flip-book, consisting typically of less than ten images, often played in a loop.

A possible additional extension for JPEG is to include support for similar animations. Meta information for timing and looping would be encoded in the syntax elements of JPEG XT by the same mechanism that already embeds the base/extension layers assembly instructions in the codestream. All frames of the animated sequence, but one, would be embedded into the JPEG codestream, hidden from legacy JPEG decoders in the same way JPEG XT hides the extension layer in the base image, namely by APP markers. Hence, animated JPEG XT would build on the technical infrastructure of JPEG XT.

To encode the metadata defining the animation sequence and its timing, the JPEG committee considers two logical choices for the syntax: The first choice would be to use the same structures and elements as in animated JPEG 2000. JPEG 2000 defines a very well developed animation system, including looping, grouping, timing and rotation of individual frames. However, even though specified in the form of an International Standard, it has not gain much popularity but just a few implementations of this format exist as of today.

The second possibility is to borrow elements from the file format of the latest MPEG standard, HEVC. This file format, known under the name HEIF (High Efficiency Image File Format) also defines syntax elements for animation and looping. It is also part of the same ISO family of formats and shares the top-level syntax with that of JPEG 2000, though the choices for individual syntax elements are a bit different. Whether HEIF will gain much attention in the future and whether it will see any real-world applications is, of course, also still to be determined.

References

[1] F. Banterle, A. Artusi, K. Debattista and A. Chalmers ”Advanced High Dynamic Range Imaging Theory and Practice”, CRC Press (AK Peters), ISBN: 978-1-56881-719-4, 275 pages, March 2011.
[2] G. Ward, “Real Pixels'', Graphics Gems II, Ed. by J. Arvo, Academic Press, 1992.
[3] OpenEXR an High Dynamic Range format Industrial Light & Magic.
[4] A. Artusi, R. Mantiuk, T. Richter, P. Hanhart, P. Korsunov, M. Agostinelli, A. Ten and T. Ebrahimi "Overview and evaluation of the JPEG XT HDR image compression standard", Journal of Real-Time Image Processing, Springer, ISSN 1861-8219, December 2015, DOI: 10.1007/s11554-015-0547-x.
[5] A. Artusi, R. Mantiuk, T. Ricther, P. Korsunov, P. Hanhart, T. Ebrahimi and M. Agostinelli "JPEG XT: A Compression Standard for HDR and WCG Images [Standards in a Nutshel]", IEEE Signal

Processing Magazine, Volume 33, issue 2, pages 118-124, ISSN 1053-5888, March 2016, DOI: 10.1109/MSP.2015.2506199.
[6] T. Richter,W. Husak, , N. Ajit, T. Arkady, P. Korshunov, T. Ebrahimi, A. Artusi, M. Agostinelli, S. Ogawa, P. Schelkens, T. Ishikawa, and T. Bruylants, “JPEG XT information technology: Scalable compression and coding of continuous-tone still images.”
[7]JPSEC: ISO/IEC 15444-8 JPEG2000 Part 8 JPEG 2000 Secure.
[8] R. Mantiuk, T. Ricther, A. Artusi “Fine-tuning JPEG-XT compression performance using large-scale objective quality testing “. Conference Proceedings IEEE International Conference on Image Processing (ICIP), Phoenix (USA); 09/2016.

Thomas Richter (richter@tik.uni-stuttgart.de)
is a researcher at the Computing Center of the University of Stuttgart, Germany. He is one of the editors of the JPEG XT standard and the chair of the Test and Quality Subgroup in JPEG.

Alessandro Artusi (artusialessandro4@gmail.com)
is a Ramon y Cajal fellow at the University of Girona, Spain. He is a member of JPEG, co-editor of the JPEG XT standard. He is the recipient of the Emerging Standards Awards from the British Standard Institute.

Touradj Ebrahimi (Touradj.Ebrahimi@epfl.ch)
is a professor at Ecole Polytechnique Fédérale de Lausanne heading its Multimedia Signal Processing group. He is also the Convenor (Chairman) of the JPEG Standardization Committee.