18

Ongoing standardization efforts

Touradj Ebrahimi, Athanassios Skodras & Peter Schelkens

18.1 Introduction

The JPEG committee has been continuing its activities on definition of international standards related to still image coding systems in order to prepare the ground for interoperable solutions, products and services, which make use of digital images. Currently, JPEG committee is pursuing three such efforts, namely, JPSearch, which deals with annotation, search and retrieval of digital images and image sequences; JPEG XR, which seeks a solution for compression of high dynamic range images, in environments such as digital photography, where there is a need for solutions with strong complexity requirements; AIC, which aims at preparing the ground for advanced image compression systems, beyond JPEG 2000 family of standards.

In the following section, each of these activities is briefly described.

18.2 JPSearch

18.2.1 Motivations

Powerful image compression and coding standards, along with rapid evolution in computer technologies, image sensors, storage devices and networking, have made digital imaging a tremendous success for consumers and businesses. As a result, a impressive growth in personal, professional and shared digital image collections has been observed in recent years. However, searching and managing these large distributed collections represents a considerable challenge. With the evolution from film to digital camera, it has become easy, fast and free to take lots of photographs. Yet, to organize the resulting digital photo shoebox, a modern variation of the traditional photo shoebox, remains cumbersome. Photo organizer applications allow for a simple classification of images to generate albums or slideshows, but often require manually annotating the whole collection, a very tedious and time-consuming process. Furthermore, because the portability of data and metadata is not guaranteed, the consumer is de facto locked into one system.

Stock photography companies, such as Getty Images or Corbis, boast tens of millions of digital photographs. These photographs are labeled and organized

2

based on professionally developed taxonomies. The tremendous cost of generating the annotation is justified and recouped by the high commercial value of the content. More recently, photo sharing websites, such as Flickr, and Picasa have gained in popularity, fueled by innovative online community tools that allow users to categorize images by use of tags, in order to allow searchers to easily find images concerning a certain topic. The resulting folksonomy consists of collaboratively generated, open-ended labels authored by the users, and sometimes submitters, of the content. However, again these systems are closed and lock up the user. Recognizing the need for an interoperable and open standard in image search and retrieval systems, JPEG has recently launched a new activity, JPSearch, also known more formally as ISO/IEC 24800. The aim of this standard is to allow different image management systems to interoperate. It is foreseen that JPSearch will enable more complete solutions and give consumers and businesses confidence in the longevity of their annotations and collection maintenance effort.

18.2.2 JPSearch overview

JPSearch aims at interoperability between devices and systems by defining the interfaces and protocols for data exchange between them, while restricting as little as possible how those devices, systems or components perform their

task. Existing search systems are implemented in a way that tightly couples many elements of the search process. JPSearch provides an abstract framework as well as a modular and flexible search architecture that allows an alignment of system design to a standard framework. In this framework, interoperability can be defined in many ways, for example between self-contained vertical image search systems that interact to provide federated search, between the different layers of image search so that these components could be supplied by different best-of-breed vendors, or at the metadata level such that different systems may add, update or query metadata for images and image collections.

In particular, JPSearch facilitates the use and reuse of metadata. A user makes a heavy investment when annotating a collection of images. With JPSearch, the portability of the metadata is guaranteed, hence allowing a user to subsequently migrate to applications or systems which best suit his/her needs. In community based image sharing systems, this portability enables the owner of an image collection to merge community metadata back into his/her own management system, hence helping to overcome the manual annotation bottleneck.

4

Similarly, JPSearch makes possible the use and reuse of ontologies to provide a common language for contexts. Indeed, searching for images always takes place in a context, either implicit or explicit. A common format for handling context allows a user to carry his/her context with him/her to different search engines. It also allows the context to be owned by the user and not by the system, hence protecting the user's privacy.

JPSearch also provides a common query language, giving search providers a reference standard to remove ambiguity in the formulation of a query, and to make searching over shared repositories consistent. The common query language also defines query management process such as relevance feedback.

Finally, by providing a solution for the carriage of image collections and associated metadata between compliant devices and systems, JPSearch enables image search and retrieval functionality across multiple repositories. Therefore, it allows leveraging the generally high cost of creating metadata.

18.2.3 JPSearch architecture and elements

A JPSearch architecture has been defined, as illustrated in Figure 18.1. This architecture is generic in the sense that most existing image search systems can be straightforwardly mapped to it.

The architecture is divided into four layers, namely:

- ❖ **User layer:** The user layer aims at the personalization of the search service.

- ❖ **Query layer:** The query layer is dealing with the formulation of the query and presentation of the results.

- ❖ **Management layer:** The management layer is handling the distribution of a search task over multiple image repositories.

- ❖ **Content layer:** The content layer includes the image repositories and the associated metadata, as well as the corresponding schema and ontology.
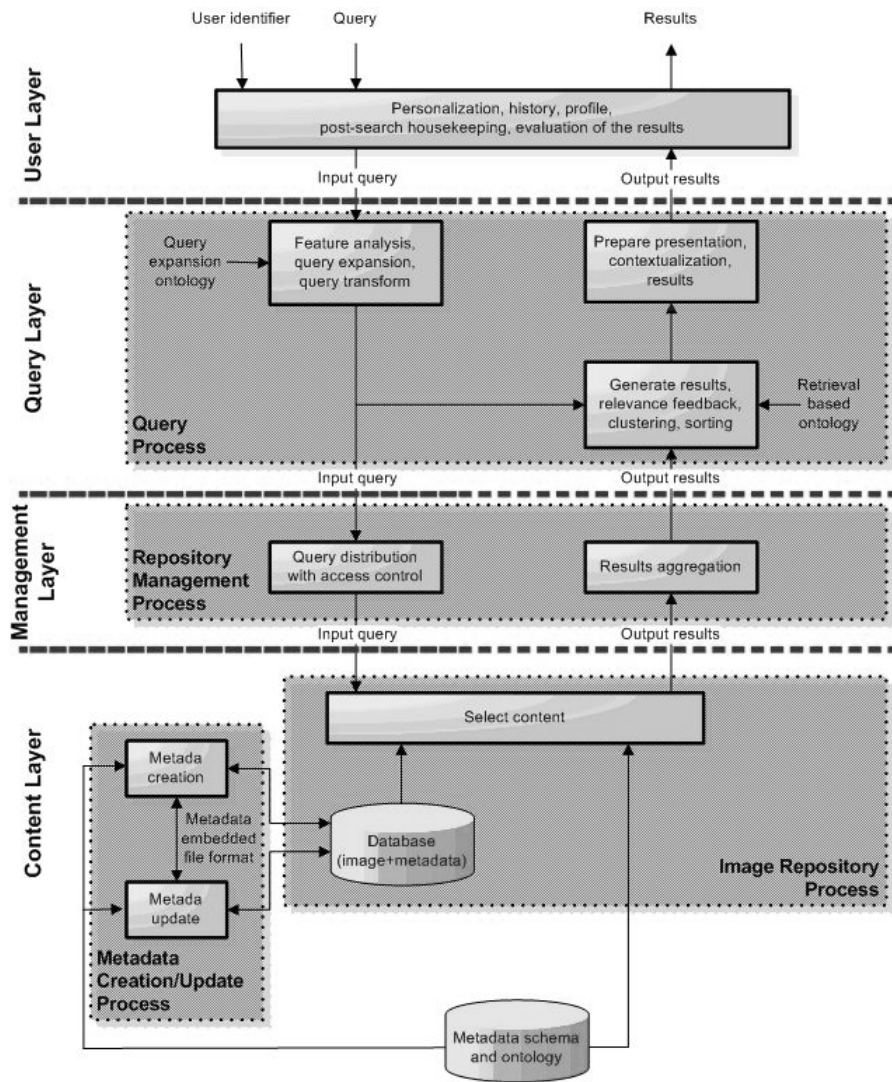
6

Figure 18.1: JPSearch architecture and layers

Note that while the query, management and content layers are in the scope of JPSearch, the user layer is outside of the scope.

In turn, four independent functional modules are involved in the architecture, as detailed below:

❖ **Query process:** The query process is part of the query layer. It aims at the efficient execution of search tasks. More specifically, it forms a machine understandable query from the user's search task which is then conveyed to the subsequent management and content layers. Conversely, it validates the search results and ranks them according to the user's criteria. Finally, the query process may also generate a new query based on relevance feedback mechanisms.

❖ **Repository management process:** The repository management process belongs to the management layer. Its purpose is to allow users to simultaneously search multiple distributed image repositories with the same query. Another purpose is to comprehensively aggregate the results returned from the image repositories.

8

❖ **Image repository process:** The image repository process is located in the content layer and supplies basic search functionalities. More specifically, this includes receiving a set of queries, executing the matching of these queries with the stored metadata, and forming a result.

❖ **Metadata creation and update process:** The metadata creation and update process is also part of the content layer. On the one hand, this process enables to build metadata using a proper schema and ontology definition. On the other hand, this process also provides with the functionality to update the metadata, e.g. by adding, replacing, removing all or part of the metadata. In particular, whenever the image content is modified, it verifies that the metadata is suitably brought up to date.

18.2.4 JPSearch specification

JPSearch is a multi-part specification. Five parts are currently envisioned, as detailed hereafter:

❖ **Part-1: Framework and system components**: This part is a type-3 technical report which provides a global view of JPSearch. In particular, the technical report reviews the traditional approaches to

image search and motivates the importance of the user in the search process. It describes a number of use cases in order to identify user needs and requirements. It then explains the overall search and management process. This leads to the introduction of a JPSearch architecture composed of 4 layers: user layer, query layer, management layer and content layer. Finally, the technical report outlines the organization of the JPSearch specification.

❖ **Part-2: Schema and ontology registration and identification**: This part standardizes a platform-independent format for the import, export and exchange of ontologies. It also defines a registry of ontologies which can be imported into a JPSearch compliant system. Finally, it standardizes basic functions to query and manipulate one or more ontologies in a repository.

❖ **Part-3: JPSearch query format**: This part provides three standardized functionalities between users and image repositories. Firstly, it allows users to express their search criteria. Secondly, it allows users to describe the aggregated return result sets for user presentation or machine consumption. Thirdly, it defines query management processes such as relevance feedback.

❖ **Part-4: Metadata embedded in image data (JPEG-1 and JPEG-2000) file format**: This part specifies image data exchange format with associated metadata to accelerate the re-use of metadata. It supports two functionalities, namely the mobility of metadata and the persistent association of metadata with image.

❖ **Part-5: Data interchange format between image repositories**: This part standardizes a format for the exchange of image collections and respective metadata between JPSearch compliant repositories. The data interchange format enables the synchronization of repositories in order to facilitate simple and fully interoperable exchange across different devices and platforms.

18.3 JPEG XR

JPEG XR – ISO/IEC 29199 – is a recent standardization effort to put forward an image compression algorithm, which provides high compression efficiency, while requiring a lower complexity when compared to JPEG 2000 compression algorithm. JPEG XR basic approach is similar to that of HD Photo image compression system proposed by Microsoft Corporation.

JPEG XR is based on a block-transform. Its coding structure, which shares many similarities with other traditional image coding techniques, is composed of the following steps: color conversion, reversible Lapped Bi-orthogonal Transform (LBT), flexible quantization, inter-block prediction, adaptive coefficient scanning, and entropy coding of transform coefficients. The distinguishing features are the LBT (Malvar, 1989)(Malvar, 1998) and advanced coding of coefficients (see Figure 18.2).

JPEG XR supports a wide range of pixel formats, including High Dynamic Range (HDR) and wide gamut formats, as well as a number of color formats including RGB and CMYK. To convert spatial domain image data to frequency domain, HD Photo uses an integer hierarchical two stage LBT which is based on a flexible concatenation of two operators: the Photo Core Transform (PCT) and the Photo Overlap Transform (POT). PCT is similar to

12

the widely used DCT and exploits spatial correlation within the block. However, it fails to exploit redundancy across block boundaries and may introduce blocking artifacts at low bit rates. To alleviate these drawbacks, POT is designed to exploit the correlation across block boundaries. Hence, POT improves compression efficiency, while at the same time reducing blocking artifacts.
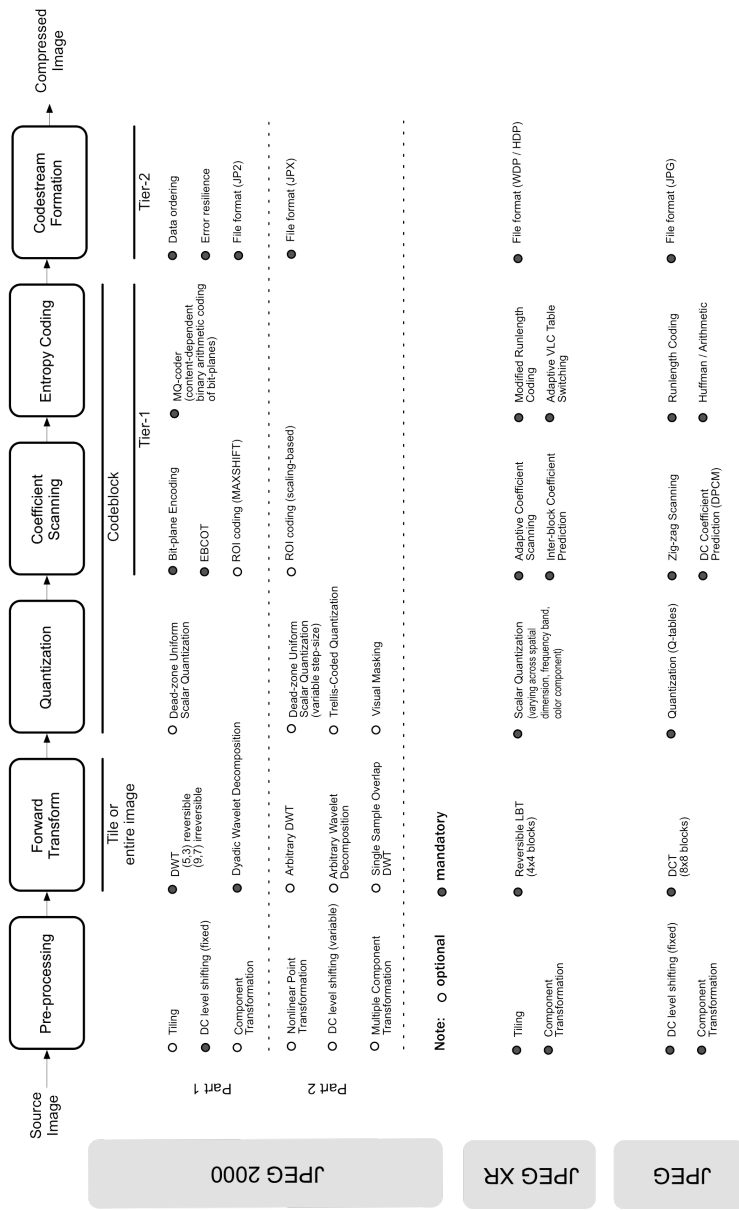
Figure 18.2: Overview of the technologies included in the different components of a JPEG, JPEG XR and JPEG 2000 encoder.

14

The transform is performed in a two-stage hierarchical structure. For the sake of simplicity, we consider the case of the luminance channel. At the first stage, a 4x4 POT is optionally applied, followed by a compulsory 4x4 PCT. The resulting 16 DC coefficients of all 4x4 blocks within a 16x16 macroblock are grouped into a single 4x4 DC block. The remaining 240 AC coefficients are referred to as the High Pass (HP) coefficients. At the second stage, the DC blocks are further processed. Another optional 4x4 POT is first performed on the DC blocks, followed by the application of a compulsory 4x4 PCT. This yields 16 new coefficients: one second stage DC coefficient and 15 second stage AC coefficients, referred to as the DC and Low Pass (LP) coefficients respectively. The DC, LP and HP bands are then quantized and coded independently. All transforms are implemented by lifting steps. The chrominance channels are processed in a similar way. Whenever POT and PCT are concatenated, the transform becomes equivalent to LBT. In order to enable the optimization of the Quantization Parameters (QP) based on the sensitivity of the human visual system and the coefficient statistics, JPEG XR uses a flexible coefficient quantization approach. Namely, QP can be varied based on the spatial location within the image (e.g. at frame, tile or macroblock level), the frequency band and the color channel.

To further improve compression efficiency, inter-block coefficient prediction is then used to remove inter-block redundancy in the quantized transform coefficients. More specifically, three levels of inter-block prediction are supported: prediction of DC sub band coefficients, prediction of LP sub band coefficients, and prediction of HP sub band coefficients. This prediction step can be compared to the Intra prediction used in H.264/AVC. Adaptive coefficient scanning is then used to convert the 2-D transform coefficients within a block into a 1D vector to be encoded. Scan patterns are adapted dynamically based on the local statistics of coded coefficients. Coefficients with higher probability of non-zero values are scanned earlier. Finally, the transform coefficients are entropy coded. For this purpose, advanced entropy coding using adaptive VLC table switching is used. A small set of fixed VLC tables are defined for each syntax element. The best table is then selected based on local coefficient statistics. The choice of table is computed based on previously coded values of the corresponding syntax element. Furthermore, HP coefficients are layered coded. More precisely, the coefficients are partitioned into two components by adaptive coefficient normalization. After the partition, the significant information is entropy coded and the remainder is signaled using fixed length codes.

16

18.4 Advanced Image Coding and Evaluation Methodologies (AIC)

The state-of-the-art in rate-distortion performance assessment has been so far the measurement of PSNR, MSE and/or MAXimum Absolute Difference (MAXAD) between the original and the decoded images for a given bit-rate in order to obtain objective measures for the reconstruction quality. To estimate though the effective visual quality, these measurements are typically combined with subjective visual tests under controlled environments (well-controlled set of test persons, well-defined viewing distances and illumination conditions …). Since the latter procedure is time consuming and thus a costly activity, and moreover, the objective measures are not guaranteeing a reasonable correlation with human visual perception, the image quality research community has massively subscribed itself to the quest for more reliable objective perceptual metrics. In this context, the JPEG committee has pre-examined several metrics that potentially contribute to an improved objective quality assessment (De Simone, 2008), as there are:

❖ The monochannel *Mean Structural SIMilarity index* (MSSIM) (Wang, 2004) that builds upon the hypothesis that the human visual system is adapted to extract structural information from the scene.

17

❖ The monochannel *Visual Information Fidelity measure in Pixel domain* (VIF-P) (Sheikh, 2006), which is based on the assumption that images of the human visual environment are all natural scenes with similar statistical properties. Gaussian scale mixture (GSM) models in the wavelet or spatial domain are use to represent the statistical properties by Natural Scene Statistics (NSS) models.

❖ The monochannel *PSNR Human Visual System Masking metric* (PSNR-HVS-M) is a modified version of the PSNR, incorporating a model of the human visual system Contrast Sensitivity Function (CSF) and contrast masking property (Ponomarenko, 2007).

❖ The multichannel *DCTune metric* (Watson, 1993) based on the computation of an error map as the arithmetic difference between DCT coefficients of the test and reference images, whereas the error DCT coefficients are weighted by their absolute visibility as a function of the DCT frequency, the display mean luminance and resolution. Contrast masking further corrects the error sensitivity.

❖ The multichannel *Weighted PSNR metric* weighs separately the contributions from the different component planes (RGB, YCbCr …).

❖ The multichannel *Weighted MSSIM metric* weighs in a similar way the contributions the different components.

Despite these metrics, investigations by the JPEG committee have shown that further efforts appear to be necessary in order to come forward with reliable and well-characterized perceptual visual metrics.

Advanced Image Coding and Evaluation Methodologies (AIC) is a new work item in JPEG 2000 with the aim to explore those new metrics and test procedures, and moreover to design the next generation image compression systems beyond those offered by JPEG 2000 family of standards. The initial goal of this activity is to assess the functionalities of current JPEG, JPEG 2000, and JPEG XR algorithms such as their compression efficiency, and other offered features, when compared to the state-of-the-art. To this end, a call for evidence for advanced image coding technologies and evaluation methodologies has been issued with responses expected to include a comparison between various functionalities (including compression efficiency) of the JPEG, JPEG 2000 and JPEG XR algorithms compared to alternative technologies, and potential candidates for AIC.

The findings of this exploratory phase will be summarized in a technical report, which will include evaluation process and conditions, as well as the

content, evaluation metrics, and results of such assessments. If there exist one or more technologies providing significantly better existing functionalities, or provide new functionalities useful for a large enough class of applications, a call for proposals will be issued in order to create a new image coding standard known as AIC. The objectives of AIC could be either offering significantly better compression efficiency (e.g, twice or more better) when compared to current JPEG, JPEG 2000, and JPEG XR standards, and/or offering advanced functionalities such as better, more efficient, or new ways to process and represent new types of images. The standardization timeframe is planned to take place in 2010 or beyond.

18.5 References

F. De Simone, D. Ticca, F. Dufaux, M. Ansorge and T. Ebrahimi, "A comparative study of color image compression standards using perceptually driven quality metrics", Proc. SPIE, Vol. 7073, Optics and Photonics 2008, Applications of Digital Image Processing, San Diego, August 2008.

F. Dufaux, M. Ansorge, T. Ebrahimi, "Overview of Jpsearch: A Standard for Image Search and Retrieval" - Content-Based Multimedia Indexing, CBMI'07, London, June 2007.

20

ISO/IEC, "JPSearch Executive White Paper", ISO/IEC JTC1/SC29/WG1 N4108, December 2006a.

ISO/IEC, "ISO/IEC PDTR 24800-1: Framework and System Components", ISO/IEC JTC1/SC29/WG1 N4083, November 2006b.

H. S. Malvar and D. H. Staelin, "The LOT: transform coding without blocking effects", *IEEE Trans. ASSP*, 37(4), 553-559, 1989.

H. S. Malvar, "Biorthogonal and nonuniform lapped transforms for transform coding with reduced blocking and ringing artifacts", *IEEE Trans. Signal Processing*, 46(4), 1043-1053 1998.

N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions", Proceedings of the Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM-07), Scottsdale, Arizona, USA, 25-26 January, 2007.

H.R. Sheikh.and A.C. Bovik, "Image information and visual quality", *IEEE Transactions on Image Processing*, vol.15, no.2,pp. 430- 444, February 2006.

Sridhar Srinivasan, Zhi Zhou, Gary J. Sullivan, Robert Rossi, Shankar Regunathan, Chengjie Tu, and Anirban Roy, "Coding of high dynamic range

images in JPEG XR / HD Photo", Proc. SPIE, Vol. 7073, Optics and Photonics 2008, Applications of Digital Image Processing, San Diego, August 2008.

Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity", IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, April 2004.

A.B. Watson, "DCTune: A Technique for visual optimization of DCT quantization matrices for individual images", Society for Information Display Digest of Technical Papers XXIV, 946-949, 1993.