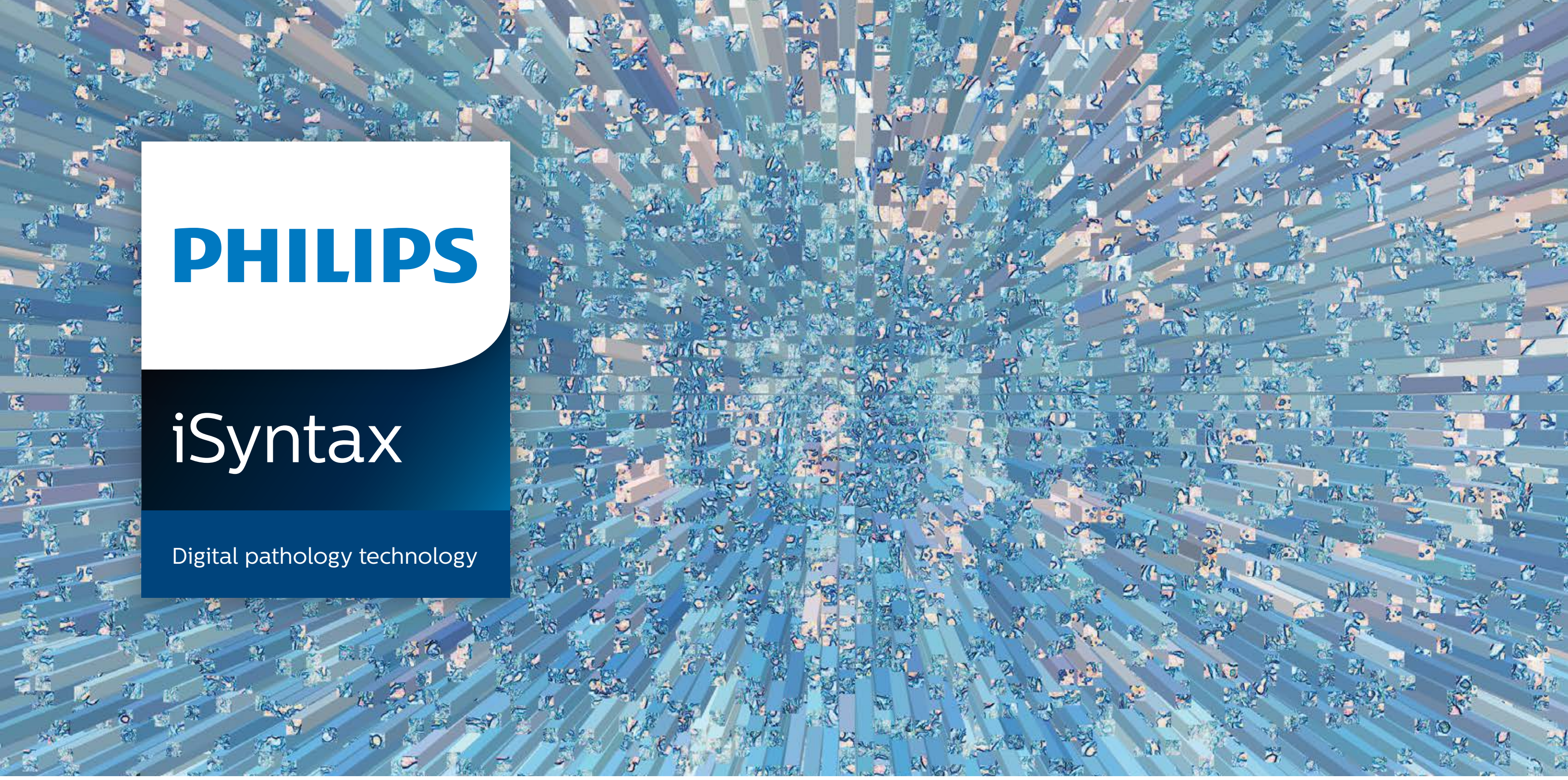

Philips' iSyntax for Digital Pathology

Dr. Bas Hulsken, CTO Philips Digital Pathology Solutions

Digital pathology requires large amounts of gigapixel images to be generated, stored and delivered with medical grade image quality and high performance, in order to deliver a seamless digital workflow. Currently available image formats do not fit the bill. JPEG 2000 certainly delivers the medical grade image quality, but is computationally too demanding to allow both fast whole slide scanning and fast viewing in web applications. JPEG is fast, supported in web browsers and allows for fast viewing, but it can not provide medical grade image quality. In this white paper we introduce Philips' iSyntax® image format for digital pathology whole slide images, which marries the medical grade image quality of JPEG 2000 with the speed of JPEG, and enables scalable cost effective image storage both on premise and in the cloud.

1 Introduction

Digital pathology Whole Slide Images (WSI) are gigapixel sized images, that need to be created in less than a minute on a fast whole slide scanner such as the Philips Ultra Fast Scanner. Additionally a suitable image format for WSI needs to support fast zooming and panning. Currently the most common image format used for storing whole slide images is the TIFF image format with the JPEG image compression method. While this satisfies the speed requirements both for image compression as well as for fast image decompression in web applications, the JPEG compression standard is a consumer grade compression method, and not optimal for medical grade images. In particular the decimation of the color resolution and the low dynamic range of 8 bit per channel and the limitation to three color channels is problematic. Additionally JPEG is not an hierarchical image compression method, meaning that it is not possible to quickly retrieve a lower resolution representation of a subregion of an image. This problem is typically overcome by storing multiple lower resolution versions of the image in the same TIFF file - often referred to as the pyramid TIFF format; this however results in a 30% larger image file at the same image quality level, or in a lower image quality at the same file size. The limitations of the now 24 year old JPEG standard are

November 2016

well known, and in the year 2000 the JPEG group created JPEG 2000 to supersede the JPEG standard. JPEG 2000 takes away most if not all limitations in JPEG that make it less suitable for medical grade images: it supports high dynamic range images (up to 32bit per channel), it does not force a color resolution decimation, and it supports multiple color channels. In addition, because it is a wavelet based compression method it is inherently a hierarchical image format which allows for fast retrieval of lower resolution versions of the original image without the need for storing the redundant pyramid of lower resolution versions of the image as required by the JPEG TIFF image format.

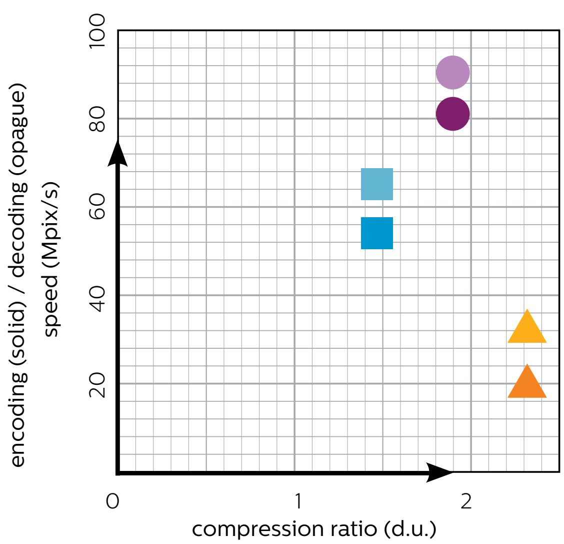

Figure 1: Lossless compression performance of iSyntax (purple circles) versus JPEG (blue squares) and JPEG 2000 (orange triangles). This graph illustrates the balance that was chosen between compression speed and compression ratio for iSyntax. iSyntax gives very good compression at speeds that allow for fast whole slide navigation in a web application.

Unfortunately despite it's clear improvements over the JPEG standard, today JPEG 2000 is a little used image compression method, even in the domain of medical imaging. The reason for this is the trade-off chosen in JPEG 2000 between compression ratio and speed: JPEG 2000 is computationally very expensive compared to the standard JPEG image compression method.[1)] And to make matters worse, for high quality images (implying a low compression ratio) required for medical applications, it does not produce smaller files than JPEG.[2)] So, while JPEG 2000 takes away most of JPEG's limitations for medical grade imaging, it's poor speed make it less suitable for digital pathology, where very large images need to be compressed in short times, and responsive navigation of large images is required, even on less powerful platforms such as web applications and mobile devices.
The Philips' iSyntax image format for pathology whole slide images was designed to combine the medical grade image quality of JPEG 2000 with the speed of JPEG, leveraging Philips' leading iSite iSyntax image representation for radiology images.[3)] The performance of iSyntax versus JPEG and JPEG 2000 is compared in Figure 1.

2 Image Compression Method

To overcome JPEG 2000's limitations while maintaining the image quality and feature set that would make a faster JPEG 2000 so suitable for medical imaging applications, iSyntax employs the same recursive wavelet transformations as JPEG 2000, but it replaces the entropy coding part by a completely new method. The entropy coding part of an image compression method is responsible for finding a more compact way to store exactly the same information. It is by definition fully reversible, meaning that the entropy coder has no influence on the image quality of the compressed image, but only on the compression ratio, and speed. JPEG 2000's EBCOT [4)] entropy coder is very efficient in terms of compression ratio, but unfortunately very slow as well. iSyntax's compression method, shown in Figure 2 uses a far simpler entropy coder, based on exploiting correlation of data in a local neighborhood combined with a fast modified Huffman coder. With this new entropy coder iSyntax delivers the same features and image quality as JPEG 2000, but with an order of magnitude faster compression and decompression. This greatly improved speed comes at the cost of an approximately 10% increase in size of the compressed image. ISyntax outperforms JPEG in terms of compression ratio, as it's wavelet hierarchy makes storing lower resolution versions of the same image superfluous. The efficient iSyntax compression method, allows for a fast HTML5 whole slide viewer that achieves video rate performance.

3 Secure on Premise and Cloud Storage

The ability to have access to your data from anywhere and on any device is increasingly relevant, also in the healthcare sector. HTML5 web applications such as the Philips IntelliSite Pathology Solution address this need by turning any web browser equiped device into a workstation. While the web application is the most visible component to the end user, in the background an equally big challenge is serving the huge amounts of whole slide image data to the application, anywhere in the world, at high speeds, and with maximum security.
To this end, iSyntax was specifically designed as a cloud storage enabled image format. It's patented data model with three levels of hierarchy is particularly suitable for efficient storage and fast retrieval on state of the art scalable and cost effective cloud storage technology, such as SAS or Solid State Disks. The weeks after sign out however, the image is less likely to be viewed, and slower, cheaper, on premise storage might suffice. Finally after the patients treatment ends, images are extremely unlikely to be accessed and might be stored on an even cheaper storage medium such as tape or in the cloud storage solutions such as Amazon S3.

over https, by allowing digital signing of the image data to ensure integrity, and by allowing at rest encryption of the image data to facilitate privacy even in the case of direct access to the storage media resulting from a physical security breach.

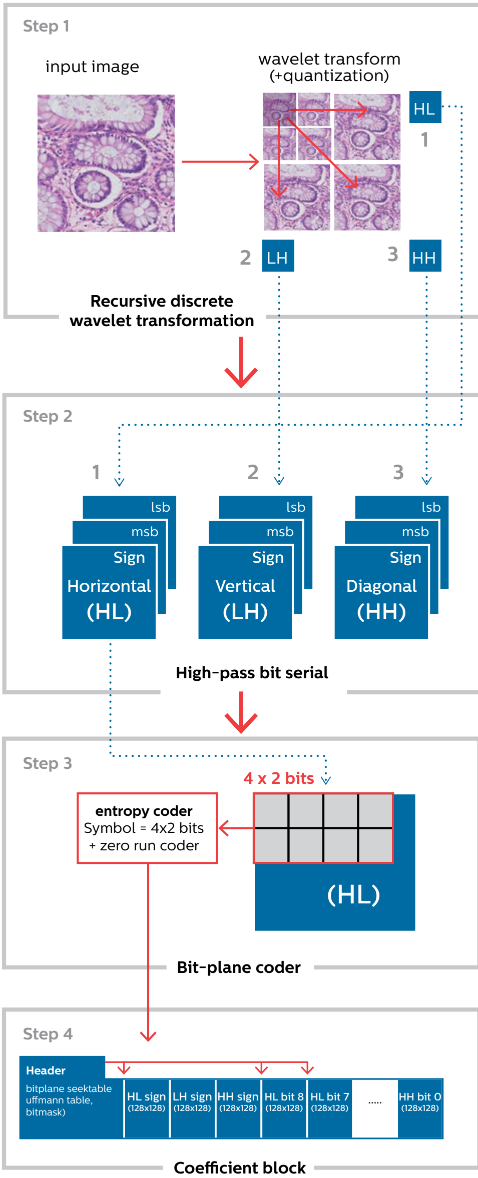

Figure 2: Schematic diagram of the iSyntax patent pending data compression method. The first step is a recursive discrete wavelet transformation of the image, the diagram shows a two step recursion. Subsequently the individual coefficient types are transformed to a bit-plane serial representation. These bit planes are then entropy coded in groups of 4x2 neighboring pixels and stored in order of significance. It's this last entropy coding part that is unique to iSyntax, the method is far less computationally complex than JPEG 2000's EBCOT method, but only 10% less efficient in terms of compression ratio.

iSyntax' speed and cloud enabled data model allow the creation of true cloud deployed solutions for digital pathology that benefit from the cloud's virtually infinite scalability and cost effectiveness. Security and privacy are key for both on premise as well as in cloud deployments of digital pathology, and iSyntax meets the highest standards by providing full in transit encryption of data

4 Storage price/performance optimization

Given the large amounts of image data generated in a digital pathology lab, it's worthwhile to design complex multi-tiered storage solutions to achieve the optimal price/performance ratio at each stage of the image data life cycle. For instance: immediate access to a whole slide image is essential while the case is not yet signed out, and this warrants expensive fast storage technology, such as SAS or Solid State Disks. The weeks after sign-out however, the image is less likely to be viewed and slower, cheaper onpremise storage might suffice. Finally after the patients treatment ends, images are extremely unlikely to be accessed and might be stored on an even cheaper storage medium such as tape or in the cloud.

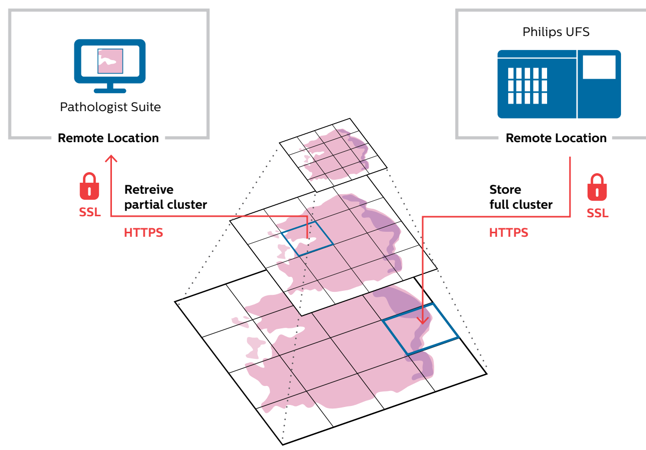

Figure 3: iSyntax' patent pending hierarchical storage model allows for storing multidimensional image data of virtually unlimited size on scalable onpremise storage systems and on cloud storage services. The first level of the threefold hierarchy of the iSyntax data model is generated by a seamless wavelet transformation of the image data. Clusters, which are regions of the wavelet transformed data spanning both space as well as wavelet scale make up the second level, finally tiles, small granular patches of image data at a single wavelet scale make up the third level. This hierarchy allows for a best of both worlds storage model; large clusters for fast streaming storage of images from the scanner and fine grained tiles within the clusters for fast random access by the viewer at any position, at any resolution. For outstanding security and reliability, the iSyntax format has built in data signing and encryption both at rest and in transit using md5 and https respectively.

The iSyntax image format's unique hierarchical data model allows for an even more granular optimization of image storage price/ performance ratio.

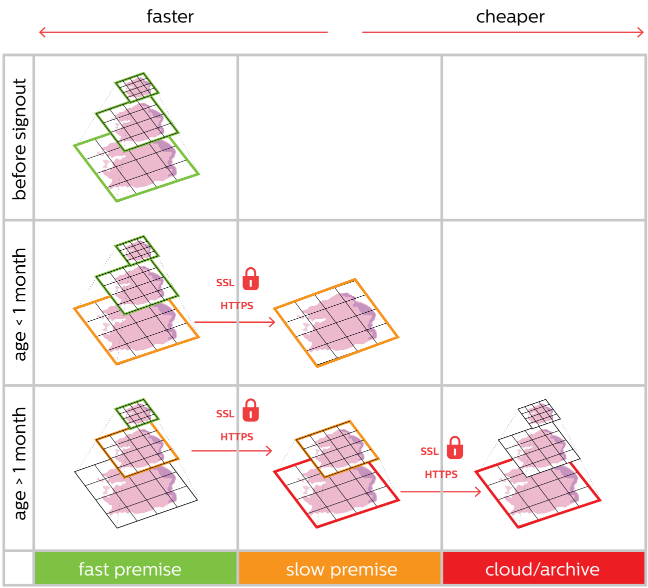

Figure 4: Example configuration of a multi-tiered image storage solution consisting of a fast on premise tier, a slower on premise tier, and finally an off premise archive or cloud storage tier. In this example all zoom levels of the image are stored on the fastest storage tier before sign out of the case. After sign out, the top zoom levels are kept on fast storage for one month, while the highest zoom level is moved to the slower and cheaper tier 2. Finally, after 1 month, only the lowest zoom level is kept on fast storage, while the medium and highest zoom levels are moved to the slow on premise tier and the archive tier respectively. All moving operations of the image data over the tiers can be performed in the background, and over the secure https protocol.

As shown in Figure 4, different resolution levels of the same image can be stored over multiple storage tiers. This allows for rapid access to the lower zoom levels of a whole slide image by storing it on fast on premise storage, and slower access only when zoom-ing in to the high resolution levels. For example, storing a 40x image with levels up to 5x on fast storage and 10x and more on slow storage, gives a factor of 64 reduction of space and cost on the fast storage tier. Any combination of magnification level and storage tiers is possible, and re-balancing the different magnification levels over the different storage tiers does not require any processing of the image data and is therefore fast and has no effect on the image quality.

5 Conclusions

By choosing a trade-off between image compression ratio and speed, the iSyntax image format is ideal for digital pathology in today's world where scalable web applications and cost effective cloud storage are must haves.

ISyntax combines JPEG 2000's medical grade features, such as greater than 8 bit dynamic range, multiple color channels and wavelet based compression, with the speed and web readiness of JPEG. This unique combination enables web applications capable of fast whole slide image viewing with medical grade image quality. ISyntax' hierarchical data model allows for smart storage systems that optimally balance storage cost versus storage speed, both on premise and in the cloud.

ISyntax' high image quality, it's unique distributed data model supporting modern scalable cloud and on-premise storage solutions, allow for efficient and scalable whole slide image analytics. A deeper analysis of iSyntax for whole slide image analytics will be presented in a follow up white paper. With iSyntax Philips IntelliSite Pathology Solution provides a future proof whole slide image management system that can cater to any lab's storage needs in a cost effective way, while maintaining medical grade image quality for diagnosis and scientific studies.

References

[1] I. Moccagatta and M. Z. Coban: Proc. SPIE 4115, Applications of Digital Image Processing XXIII (2000) 521.

[2] F. Ebrahimi, M. Chamik and S. Winkler: Proc. SPIE 5558, Applications of Digital Image Pro- cessing XXVII (2004) 300.

[3] J. Huffman: MEDICAMUNDI 51/1 (2007) 36.

[4] D. Taubman: IEEE Trans. Image Proc. 9(7) (2000) 1158.

4522 207 34721 - November, 2016

PHILIPS